United States Department of Agriculture

National Agricultural Statistics Service

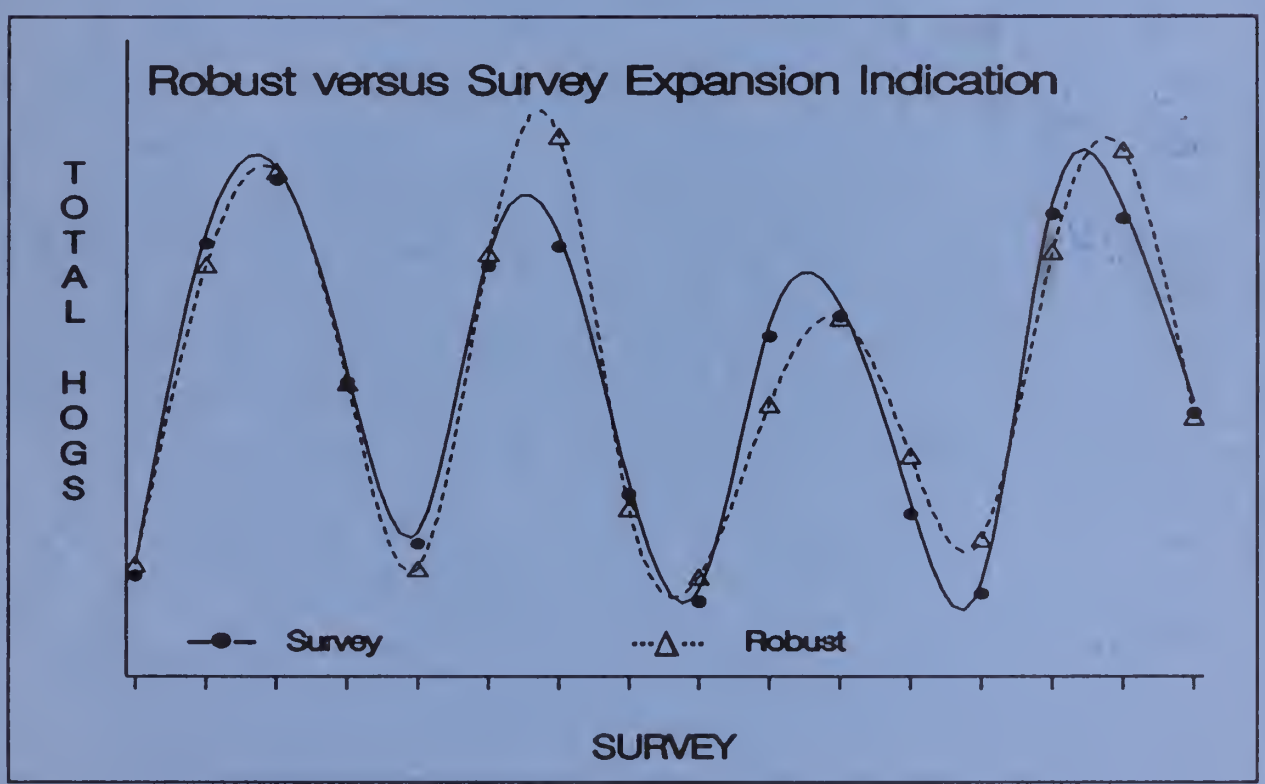
Research and Applications Division

SRB Research Report Number SRB-92-02

March 1992

CHARACTERISTICS OF DIRECTLY EXPANDED HOG DATA OUTLIERS

Scot Rumburg



Characteristics of Directly Expanded Hog Data Outliers, by Scot Rumburg, Research and Applications Division, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C. 20250, Research Report No. SRB 92-02, March 1992.

ABSTRACT

The National Agricultural Statistics Service (NASS) conducts quarterly multiple frame (MF) hog inventory surveys, using both a list and an area sampling frame. The MF survey direct expansion (DE) estimator for hogs is used as an indication by the National Board. A Robust Estimator (RE) is also currently being used by the National Board to augment the survey direct expansion total hog estimate. The first component of the RE is a DE non-outlier component. The second component of the RE is an average outlier component determined using data from several surveys. This report examines characteristics of those outlier records that are fundamental to computing the RE's outlier component.

Individual characteristics for outlier non-overlap (NOL) records for five states were studied. Three major causes of NOL outlier occurrence were found. They include increased expansion factors due to subsampling in follow-on surveys, the transitory and varying nature of hog production, and the location of hog operations on land with little or no agriculture. Individual list and NOL hog operation often produced more than one outlier within a frame-year.

Results indicate that at the state level, some survey frame-year outlier totals and number of outliers occurring differ significantly. (The frame-year runs from June through May.) At the national level, no significant effects were found. Also, distributional differences for outliers were found between states. This is due in large part to the outlier cutoff values assigned each state.

With no trends in outliers at the national level, the Robust Estimator should continue to be used without modification. However, the presence of a frame-year effect in outlier totals and outlier number of occurrences at the state level not only justifies the need for the RE but implies a modified RE might be investigated to help in making state level estimates. State level outlier cutoff values should be investigated in an attempt to locate optimum values for the RE's ability to measure state and national board estimates.

Keywords: robust estimator, outlier component, outlier cutoff value, non-overlap, frame-year, expansion factor

Washington DC March 1992

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The author would like to thank Jerry Thorson, Don Allen, Bill Iwig, Fred Warren, Gary Keough, and Charles Perry for taking the time to answer numerous questions and for the help they provided.

Additionally, further appreciation is extended to Charles Perry and Gary Keough for their introductory work in this area.

SUMMARY

The National Agricultural Statistics Service conducts quarterly multiple frame (MF) surveys to estimate total hog numbers. Both the list and area-frames are developed and fixed before the June survey and used for four consecutive surveys. The June *base* survey with the September, December, and March *follow-on* surveys define a frame-year. Homogeneous list strata and area strata are defined to reduce variance and sample size.

The total hog direct expansion (DE) estimate for a survey is the sum of all usable DE records. DE records are the product of actual operation hog numbers and an expansion factor. Outlier DE records occur when a record expands beyond the cutoff value specified for a state. Records which are in the area-frame sample but not on the list frame are denoted as non-overlap (NOL) records. These NOL records often have very large expansion factors as compared to list records and account for many of the large outliers in a survey.

A Robust Estimator (RE) is currently being used with the DE estimator to help the National Board in making hog estimates. The RE treats non-outlier data exactly as the DE estimator would, but calculates an average value using the combined outlier data from several surveys for the outlier component. This lessens the impact of unusually large individual outlier values occurring on a given survey.

Data were extracted from quarterly survey data sets from June 1987 through December 1990 for five selected states (Colorado, Georgia, Idaho, Michigan and Illinois). Individual and state level outlier analyses were performed on the data. A comparative analysis using the combined 5-state data and national 48-state summary data was also conducted.

Individual NOL outliers were tracked through frame-years to study causes of outlier occurrence. The reasons for NOL outlier occurrence are varied. For many NOL outliers, the follow-on subsampling scheme increased an already large expansion factor, creating or enlarging NOL outlier records. Over one-fourth of all NOL operations that produced an outlier were found to do so as a direct result of increased expansion factors with little change in hog production characteristics.

A second cause of NOL outliers was the transitory and varying nature of hog production. One-fourth of the NOL operations producing outliers lacked hogs in at least one survey yet produced enough hogs in a later survey during the same frame-year to create an outlier. Over forty percent had more than three times as many hogs during one survey, compared to the other surveys during the frame-year.

A third source of outlier origination was the area-frame stratification of land by agricultural intensity that does not indicate well the potential for hog production. Nearly thirty percent of NOL operations producing outliers were found on stratum 30 or greater.

It was also found that several operations produced more than one outlier during the frame-year. It is important to remember that a record is NOL because the operation is not found on the list. Improved list building and maintenance would reduce the number of NOL operations, and also both outlier and non-outlier NOL records.

Analysis of variance (ANOVA) was performed at the state, 5-state and national level seeking frame-year and quarterly (seasonal) trends. At the state level the number of outliers occurring, outlier magnitude per occurrence and outlier total per survey were the variables of interest. ANOVA tests of statistical significance revealed the presence of a frame-year effect for the outlier totals per survey and number of outliers occurring per survey. This effect may be due to procedural changes that occur between frame-years.

These changes include NOL operations being placed on the list-frame, operations being rotated off the area-frame or differences in samples. It also may indicate changes within hog production at the state level. No changes or trends were found in the magnitude of individual outliers over frame-years or quarters for any state. At the combined five-state "regional" level and at the national level, outlier totals per survey and outlier occurrences per survey were investigated. At the regional level, ANOVA tests found a frame-year effect in outlier totals per survey and no quarterly differences. ANOVA of national outlier summary data was performed over the same 15 surveys for 48 states using the summarized hog data from those surveys. Nationally, neither a frame-year nor a quarter effect was detected for outlier survey totals or number of outliers occurring per survey.

Summaries of outlier data from each state show that outlier distributions across states differ, even for states with similar hog production. The specified state cutoff values for outliers (values which dictate at what magnitude expanded records are defined to be outliers) influence directly the number of outliers and the proportion of list and NOL outliers for each state. State cutoff values for outliers appear inconsistent and it is largely these inconsistencies which create the differing outlier distributions.

Since the Robust Estimator discounts the impact of individual outliers or groups of outliers in a single survey and uses data from several previous surveys, it produces less variable estimates than the survey direct expansion estimate. This is especially true when frame-year effects are seen on outlier totals. Frame-year effects on outlier number of occurrences however, may show the need for a modified RE if this effect is due to production differences across frame-years. Thus with no trends in outliers at the national level, the RE should continue to be used without modification. At the state level, the presence of a frame-year effect in outlier number of occurrences implies a modified RE which accounts for this effect might be investigated. State level cutoffs should be investigated in an attempt to locate values which provide consistency across similarly producing states and to optimize the RE's ability to measure National Board estimates.

It is recommended that:

1. At the national level, up to 15 quarters of outlier data (the most this report can justify) should be used to compute the second component of the Robust Estimator.
2. Investigation continue at the state level comparing the simple Robust Estimator with a modified Robust Estimator which recognizes the frame-year effect for outlier number of occurrences and their effectiveness at setting state level estimates.
3. Sample design be examined to seek ways to reduce the number of NOL outliers by reducing follow-on expansion factors or the number of NOL records. Follow-on NOL expansion factor reduction might be accomplished by asking the peak number of hogs an operation expects over the course of the frame-year, asked during the June survey, or by using the most recent previous survey data, if available, to stratify an operation in the current survey. The number of NOL records could be reduced through increased list building and maintenance.
4. Investigation be instituted to study the impact that different state cutoff values and number of surveys used have on improving the Robust Estimator. Changes in cutoff values could result in an increase in data processing if both old and new cutoff values are maintained.

INTRODUCTION

The Robust Estimator (RE) for multiple-frame indications has been used by the National Agricultural Statistics Services's (NASS) Agricultural Statistics Board in helping to set the national hog estimates since December 1989 and has proved to be beneficial. (For origins of the RE see [Thomas, et al.]). This estimator is designed to lessen the impact of large outlier* records from a single quarterly survey by averaging outlier values over several surveys. However, specific characteristics other than magnitude of the outliers which are being averaged have not been closely examined.

The RE is currently being considered for other commodities as well as helping to set state-level estimates. It is at the state level that the RE would prove most helpful since it is at this level that outliers have their largest influence on a survey estimate.

A first objective of this report was to investigate potential causes of outlier occurrence for operations which are sampled on the area-frame but are not contained on the NASS list frame (NOL records). Some of these outliers expand to extraordinary size and can be a substantial proportion of a state's total hog estimate. If changes could be instituted to lessen the number of these outliers that occur, or at least lessen the magnitude of these outliers when they do occur, then any survey estimator would be improved. However, the improvements arising from these changes must be evaluated against any costs incurred. This could include increases in respondent burden, funding, labor-hours or computer time.

A second objective was to investigate characteristics of outlier records as a group at state, regional and national aggregate levels. Outliers from five selected states were studied. At the state level the variables of interest were outlier survey totals, outlier number of occurrences and outlier magnitude per occurrence. At the regional and national level variables of interest were outlier survey total and outlier number of occurrence. Effects (trends) were sought for these variables within survey frame-years and seasonally (quarterly) by survey. If the outlier component for the RE is calculated using a simple average (as is currently done) then it is assumed that all outliers are identically distributed. If however, outlier trends exist within survey frame-years or seasonally by survey, better estimators could be found which would compensate for these factors while still maintaining robustness to large outliers. A frame-year effect might imply that outliers are influenced by the sampling scheme or year to year list maintenance. A quarter effect might imply that outliers are affected by the seasonal influences known to occur in hog production.

A third objective of this report was to investigate how the composition of outlier distributions differ between states. Many states' hog production characteristics have changed since state cutoff values for outliers were begun many years ago. Since the largest outliers are nearly always NOL records, these cutoff values help govern the percentage makeup of the list versus NOL outliers.

---

* For purposes of this report, an "outlier" is any record whose hog total, when expanded, exceeds the state specified outlier cutoff for hogs. These records are considered to be unusual or influential. (This definition equates to a definition provided by [Keough & Perry, 1991]). Additionally, since the only statistical variable of interest for a record is the operation's total number of hogs, the terms "record" and "operation" will be used interchangeably.

METHODS

Survey Procedures

Direct Expansion Estimate for Hog Inventory.

NASS conducts surveys quarterly to estimate, among other commodities, total hog numbers. These surveys employ a multiple frame (MF) technique. One sample is drawn from a registry of farm operations known as the list-frame. The list-frame is not a complete frame of all agricultural operations. A second independent sample is drawn from the area-frame which encompasses the 48 contiguous states. All area within the 48 states has a positive probability of selection and therefore, the area-frame is a complete frame. Any record which is sampled on the area-frame but is not found on the list-frame is called non-overlap (NOL). The NOL records help to measure the incompleteness of the list-frame.

The list-frame is developed before the base June survey - the start of the frame-year - and is used in three subsequent follow-on quarterly surveys (September, December and March). The area-frame for each state is developed on a rotational basis about every fifteen years.

To create homogeneity and reduce variance, both area and list-frame records are stratified prior to sampling. List-frame records are stratified using a priority scheme. Data collected on livestock, crop acreage, and grain storage capacity from previous surveys and outside sources are used. (This information is called control data.) Each list record is assigned to a stratum prior to the June survey. Area-frame sampling units are called segments. They typically range in size from 0.25 to 1.0 square mile and are stratified by the amount and similarity of agricultural intensity when the frame is created for the state. This stratification can be classified into five general groups. Strata 11-19 are intense agriculture (usually 50% or more), 20-29 are light agriculture (15% to 50%), 30-39 are agri-urban areas (20 or more dwellings per square mile), 40-49 are range land (less than 15% agriculture) and 50^{+} are non-agriculture and water.

Both list and NOL records are expanded proportional to the size of the population each represents and inversely proportional to the number sampled from that population. These expansion factors are calculated for each record and the product of the expansion factor and the variable of interest (ie., total number of hogs owned by the operation) represents a direct expansion (DE) for that record. Typical list expansion factors range from 1 to 80 while typical NOL expansion factors range from 200 to 1000. The sum of all list and NOL DE records in a state produce a state level MF estimate and the sum of all state level MF estimates produce a MF DE estimate for the national total hog inventory.

The DE hog estimate for an area record is a weighted estimate. This means that the record, once expanded, is prorated back to the sampled area segment for an operation. Thus a DE area record will have an area adjustment weight assigned which is the ratio of the operation's land inside the segment to the operation's total land (the sum of an operation's land both inside and outside the segment). In the case of refusals and inaccesables in June, the amount of land outside the segment is unknown and often is set equal, or nearly so, to the amount of land the operation maintains within the segment. This can occur even though the operator may own considerable land outside the segment. This results in a smaller downward adjustment than should be and is often coupled with a large expansion value in follow-on surveys. Though any improper area adjustment can for an area operation can result in outlier creation, this non-sampling error is not easily studied and analysis of area adjustment problems will be limited to the NOL operations

described above which refused or were inaccessible in June but later produced a positive DE record for hogs. For more information on this problem see [Pafford, 1990].

If a large expansion factor occurs with a large hog operation an extraordinarily large DE record will be created. If these records expand beyond a certain value they will be labeled as outliers. (This is explained further in the following Outlier Identification section.) Though these large DE records are statistically justifiable - they represent unusually large operations for their stratum - they are a rare event. Their absorption into a single survey causes that particular survey indication to be inflated, *overrepresenting* the rare population of that farm type for that stratum and survey). Until another of this farm type is selected, survey indications will be deflated modestly, underrepresenting the rare population of that farm type for that stratum and those surveys). On average over time, this rare population will be represented correctly. Therefore, there is justification of and attraction for the Robust Estimator which uses the information from several surveys in estimating hog totals. (For more information on NASS multiple frame estimators see [Nealon, 1984].)

Outlier Identification.

Any record which expands beyond a state's outlier cutoff value is considered an outlier. (The exception is very large list hog operations. These operations are sampled with certainty in each quarterly survey and are not counted as outliers if they exceed the cutoff value.) These cutoff values which define an outlier were prescribed to each state over a decade ago based on a percentage of its total hog inventory. Large hog producing states have a higher cutoff value than smaller hog producing states. These cutoffs are a major factor in composing the outlier distributions for all states, and also nationally. For a more detailed discussion of state specified outlier cutoffs and detection see [Keough & Perry, 1991]. Since a DE record is categorized as either an outlier or a non-outlier for hogs, the DE estimate for state and national total hog inventory can be broken down into an outlier total and a non-outlier total component. The sum of these two components is the total hog inventory similar to the list and NOL component described above.

Outlier DE records from both the list-frame and the area-frame are always present in hog surveys at the national level. Larger hog producing states tend to have more outliers than smaller hog producing states and some of the smallest hog states often have no outliers for a given survey. The largest DE records tend to be NOL records since they have much larger expansion values. Many things influence the size and number of outliers. These can include the quality of the list and area-frames, the quality of control data, the sampling scheme, variability of operations, variability of the market, economics, weather and more.

Follow-on Sampling and Multiple Outlier Occurrences.

An operation sampled in the June base survey is often resampled during a frame-year in follow-on surveys. It is conceivable that a NOL operation could be sampled in each of the four quarterly surveys and a list operation could be sampled in as many as three surveys. Any operation sampled during any survey could potentially generate an outlier record.

The sampling scheme for follow-on surveys used during the time frame analyzed assures that many records sampled in June will be resampled. For NOL records all operations found to have hogs in June are nearly always resampled in the September follow-on survey and, if no subsampling occurs, expansion values will not change. Subsampling of the June sample always occurs in the December and March surveys and occasionally in the

September survey. This is done to reduce respondent burden on operations which are assumed to have few if any hogs and thus would contribute little to the total estimate. Prior to follow-on sampling, NOL records are stratified by the number of hogs they reported during the June survey. If a subsample is selected, more samples are drawn from strata which contain the larger NOL hog operations. NOL records with zero hogs and area records with no agriculture (non-ag) are sampled at a low rate. (See Appendix A for a more complete description of the NOL follow-on sampling procedure used during this time-frame.) Subsampling creates increased expansion factors for NOL operations selected since a single sampling unit must represent a larger portion of the target population. Expansion values for records selected for subsampling generally increase from two to five times, depending on the size of the follow-on subsample and where the NOL operation is stratified. The March survey nearly always samples the same respondents as the December (sub)sample and expansion factors are usually the same for these two surveys. List operations are resampled in follow-on surveys based on their replicate code. This code is used to maintain continuity of the hog series estimate while reducing respondent burden. The sample design rotates replicates from survey to survey, keeping a 40% overlap of the previous quarter's sample.

Alternatively, there are reasons why a record is not resampled in a follow-on survey, even if it were positive for hogs in June. All states are currently allocating 60% of their NOL records found in June for follow-on sampling in the quarterly agricultural surveys while the remaining 40% of NOL respondents are designated for other surveys. However, this allocation was phased in over time beginning in June of 1987 and completed in June 1990. (Colorado and Idaho began in 1987, Georgia in 1988, Michigan in 1989 and Illinois in 1990.) If a record is assigned to the 40% it is ineligible for resampling in quarterly agricultural surveys for that frame-year. (Additionally, any NOL record allocated to the 60% will automatically have an increase of 1.67 times its June expansion factor and any additional subsampling expansion factor increase for all follow-on surveys.) It is difficult to address the impact of the 60/40 split on follow-on expansion factors since the first, second and third largest hog producers in the five-states sampled entered into the program in the last, second to last, and third to last year respectively of the data set. Also, a NOL record may not be resampled in March if it was not selected for the December subsample. List records can be precluded from follow-on sampling if they are rotated out of the sample prior to the next survey. This can happen after any quarterly survey depending on the record's replicate code.

The Robust Estimator.

The Robust Estimator (RE), as mentioned previously, uses the sum of two components to produce an unbiased estimate for total hogs, similar to the DE total hog estimate. A comparison of the DE estimate for total hogs and the RE for total hogs is shown below. The DE estimator is shown as the sum of the outlier and non-outlier portion for comparison purposes. The first component of the RE is the non-outlier total hog DE population estimate. The second component is an *average* outlier component using both past and current outlier DE survey information. At present all quarterly surveys from March 1988 to present are being used. The RE's second component ($\overline{DE_{OTRL}}$) spreads an outlier over multiple surveys. This allows a larger group of outliers to be used to compute the second component of the RE while lessening the influence of any one outlier on a single survey. Questions about the composition of the RE's outlier component represent

two of the three objectives of this report: analysis of individual outliers and analysis of state level outlier composition.

The Direct Expansion Estimator for Total Hogs

$$DE_{TOTAL} = DE_{NON\text{-}OTLR} + DE_{OTLR}$$

where

$DE_{NON\text{-}OTLR}$ = *DE total for all non-outlier DE records*

DE_{OTLR} = *DE total for all outlier DE records*

The Robust Estimator for Total Hogs

$$RE = DE_{NON\text{-}OTLR} + \overline{DE_{OTLR}}$$

where

$\overline{DE_{OTLR}}$ = *An average survey outlier total, calculated over several surveys*

Selection of Representative States And Survey Years

It was desired to look at a cross section of outliers which produce the RE's second component. Five somewhat diverse states were selected: Colorado, Georgia, Idaho, Illinois, and Michigan. These states rank 22nd, 13th, 33rd, 2nd and 11th respectively, based on 1990 end-of-year hog inventory numbers. Overall, the five states have an average rank of 16.2 and account for 15.4% of total national hog inventory. A listing of the ranking of all states and their end-of-year 1990 total hog inventory can be found in Appendix B. Average survey estimates of hog production for the five states, using quarterly data from June 1987 through December 1990, are shown below.

	COLORADO	GEORGIA	IDAHO	ILLINOIS	MICHIGAN
AVERAGE SURVEY DE HOG TOTAL	(14 Surveys) 238,733	(15 Surveys) 1,244,056	(13 Surveys) 89,561	(15 Surveys) 5,444,381	(14 Surveys) 1,234,773

For the 10 largest hog producing states more than 25 years of quarterly hog survey data were available. Though Georgia ranks 13th in 1990 hog inventory it is considered a top 10 hog producing state as is Illinois. Quarterly surveys for all 48 contiguous states, including three of the ones selected to be studied, were begun in March of 1988. Prior to that all 48 states were only sampled semiannually in June and December. Thus, it was decided to begin collection of data from June 1987 to include an entire frame-year for Georgia and Illinois. Quarterly survey data were retrieved through the December 1990 survey. Problems were encountered with acquisition of the March 1988 data set for Idaho and it has only two survey data sets for the 1987 frame-year. Thus for the 1987 survey frame-year, Illinois and Georgia have data for all four quarterly surveys; Colorado and Michigan have data for June and December 1987, and March 1988 surveys; and Idaho has data for June and December 1987 surveys. All subsequent quarterly data sets through December 1990 for the five states were recovered without problem. A brief summary of data acquisition, calculation of estimates and quality of reproduced data can be found in Appendix C. For additional information on acquisition, reproduction and summarization of hog data, a guide is available from the Estimates Research Section [Rumburg, 1991].

It is worth noting that Georgia was selected partly because of the extraordinarily large outliers which it exhibited during the 1989 frame-year. These outliers have an influence

at the state level but are representative of the magnitude of outliers (at least NOL outliers) that can and do occur with any survey.

Outlier Analysis

Since NOL outliers are larger and more influential than list outliers, they were selected for study in an attempt to find any causes which could potentially be corrected. Any NOL record which expanded beyond the state outlier cutoff value was tracked through the frame-year in which it occurred, for each of the five states. NOL outlier records were then categorized in two tables. The tables attempt to categorize operations and outliers by potential causes which were found.

Analysis of variance (ANOVA) tests for statistical significance of specific trends were performed at three levels. (A description of ANOVA is provided in Appendix D.) The first level was individual outliers combined to the state level, the second level included all individual outliers from all five states combined to form a pseudo-regional level, and the third level was national 48-state summary data. The model for all three levels fit a frame-year, quarter and combined frame-year and quarter effect.

The state level is the most homogeneous grouping since outlier magnitude and distribution composition is a function of the state outlier cutoff value. Unfortunately, this group also represents a more localized set and individual outliers are afforded much more influence than at either the regional or national level. ANOVA was performed at the state level on the number of outliers occurring, outlier size per occurrence, and outlier total per survey.

The regional level ANOVA was performed on the combined five-state set of individual outliers. ANOVA was performed on outlier totals and number of outliers occurring per survey. Regional five-state analysis was restricted to only surveys in which all five states were represented (n=13 surveys).

At the national 48-state level, outlier summary totals generated from the national level quarterly agricultural surveys were analyzed. Again, frame-year and quarter effects were sought for outlier totals and number of outliers occurring per survey. National analysis was limited to surveys where all 48 states were represented (excludes September 1987). The December 1987 national summary statistics did not provide number of outliers occurring, so only n=13 surveys were available for the outlier occurrence per survey ANOVA. The outlier total per survey ANOVA used n=14 surveys.

Lastly, analysis was performed to find the state level composition of outliers. Summary statistics for outliers were categorized by list and NOL for each state to compare outlier distributions across states and evaluate the effect of the outlier cutoff value.

RESULTS

Individual NOL Outlier Results

Tracking of NOL Outlier Operations Within Frame-Years.

The five states over four frame-years (and 13, 14, or 15 surveys depending on the state) produced a total of 142 individual NOL outlier records from 76 operations. The tracking of NOL operations that produced outliers within a frame-year provided three potential causes for NOL outlier occurrence. All NOL operations which produced an outlier for the five states over the four frame-years can be found in Appendix E with appropriate survey responses and completion codes.

One cause of outlier generation for NOL operations was increased expansion factors. These increases were the result of the 60/40 split and the ensuing follow-on subsampling scheme for surveys as discussed in the Methods section. Forty-five NOL operations produced at least one outlier and also had increased expansion factors in either the September, December and/or March follow-on surveys. This represented 59% of the seventy-six total NOL operations producing outliers. Of those forty-five NOL operations, nine produced outliers due primarily to the 60/40 split, nine produced outliers due primarily to follow-on subsampling, four produced outliers primarily due to a large increase in hog numbers, seven were a combination of two or more of the above, and the remaining 16 operations produced outliers in June prior to any expansion factor increase. In fact, in at least twenty operations - nine affected by the 60/40 split, nine affected by subsampling and two affected by both the 60/40 split and subsampling - the outlier was produced as a direct result of expansion factor increases. Thus, over one-fourth of all NOL operations producing an outlier did so due to increased expansion factors with little change in hog production during the frame-year. This number would have been higher had all five states been in the 60/40 split program throughout the 15 surveys sampled. Though these results are somewhat subjective, the twenty-eight NOL operations which produced outliers only in follow-on surveys after expansion factor increases, shows that these increases are in part responsible for the production of outliers.

Expansion factor increases for the December/March subsample (where the vast majority of non-60/40 subsampling was done) ranged from a minimum of 1.67 (the 60/40 expansion implying no subsampling was done beyond the 60/40 split) to a maximum of 35 times the original June expansion factor. The average expansion increase for those records which were sampled in both the June and the December/March surveys was 2.91 times the original June expansion factor. These expansion factor increases, in general, represent the sampling interval for the follow-on sample and, with the 60/40 split now employed in all states, will always be present in any follow-on survey for all operations sampled. It is unclear how these increased expansion factors, which result from subsampling, affect the RE. Clearly, larger subsamples would generate a less variable outlier component for the estimator but again, the value of sample increases must be weighed against increased costs and respondent burden. In lieu of increased subsampling, perhaps distributing the follow-on subsample based on the maximum number of hogs which an operation expects to have over the next twelve months (asked during the June survey) might provide a better criterion than the present use of the number of hogs an operation has in June. Additionally, stratifying on data from the most recently completed quarterly survey (instead of strictly from the June survey) could also prove helpful.

Another factor seen to produce NOL outliers is the transitory nature of hog production. This transient nature can be divided into two areas - variability in the number of hogs being maintained at any given time and variability in the placement of hog production facilities with respect to overall agricultural land usage. These two factors often work in tandem to produce hog outlier records.

The first variability problem, variability in hog numbers between June and follow-on sampling, makes follow-on stratification difficult. For the seventy-six NOL operations producing outliers, 19 (25%) lacked hogs during one of four quarters surveyed and of those 19, eight (11% overall) lacked hogs in June. Yet, during another survey within the frame-year, all those operations produced enough hogs to generate an outlier record. Of

the remaining 57 operations: twelve had over three times as many hogs during their peak hog number survey (non-expanded count) as they did during their low hog number survey, eleven were sampled only once in June and were then allocated to the 40% split for other surveys, and two were sampled once in June and then went out of business. Thus of the 76 operations, 31 (41%) had more than three times as many hogs in one survey as compared to another survey in that frame-year. It is apparent from this analysis that stratifying on the June hog number does not wholly describe an operation's production characteristics across a frame year. Again, stratifying by the peak number of hogs expected over the frame-year, as mentioned above, may help improve the operation description.

The second variability problem, the variability in the placement of hog production facilities with respect to agricultural usage of the land, results in initial large expansion values. The area-frame is stratified by agricultural intensity of land usage and sampled proportionally to this intensity. Therefore, large expansion values occur in less intense agricultural areas. However, agricultural land usage and hog production are not necessarily associated. Of the seventy-six NOL operations which produced outliers, seventeen (29%) were located on stratum 30 or greater. The establishment of hog operations on low and non-agricultural intense areas often lead to large expansion factors and large DE hog records.

Two examples of NOL operations showing variability in hog numbers and location of facilities on light intensity agricultural land are shown below. The first operation shown in Figure 1a had zero hogs in June but later entered hog production. This operation was located in Michigan on a 31 stratum (agri-urban, more than 20 dwellings per square mile) where hog production is not commonly found. The small June sampling rate for a low intensity agricultural stratum results in a large initial expansion. The lack of hogs places the operation in a subsampled follow-on stratum (with a sampling interval near two) for December/March resulting in a near doubling of the expansion factor. The result is a large March DE hog record. Michigan's cutoff value for an outlier is 15,000.

FIGURE 1a.

SURVEY	EXPANSION FACTOR	HOG TOTAL	DE HOG TOTAL
JUNE 88	299.30	0	0
SEPT 88	299.30	111	33,048
DEC 88	561.19	0	0
MAR 89	561.19	152	84,854

AN NOL operation with varying hog numbers across the '88 frame-year

The second operation shown in Figure 1b was located in Georgia (outlier cutoff value of 25,000) on a 40 stratum (less than 15% cultivation), where again, such activity is not usually expected. During the frame-year this operation's hog total increased nearly 10 fold. The expansion factor increase because of the 60/40 split in the September, December and March subsample only increases an already large DE record. Though this case is extraordinary in the magnitude of the hog totals, several NOL operations had a much larger percentage increase between surveys (see appendix E).

FIGURE 1b.

SURVEY	EXPANSION FACTOR	HOG TOTAL	DE HOG TOTAL
JUNE 89	185.90	654	59,868
SEPT 89	309.83	3,157	481,657
DEC 89	309.83	4,680	714,017
MAR 90	309.83	5,986	913,271

An NOL operation with a nearly 10-fold increase in hogs within the '88 frame-year

Two Classification Schemes for NOL Outliers.

Two classification tables were produced, based on NOL operation characteristics and the outliers they produced. These tables provide further clarification of some causes of NOL outliers. The first classification was performed on the 63 NOL operations sampled more than once during the frame-year. (Eleven of the 76 operations were sampled only during the June survey and two were out of business for the remaining three follow-on surveys.) These 63 operations produced 129 outliers which were categorized by hog variability, expansion factor increases, and June agricultural land-use strata.

Figure 2a below provides an overview of conditional and marginal distributions for the three factors found to produce NOL outliers. It shows how the 129 NOL outliers produced are distributed with respect to the three potential outlier causes. Forty-three percent of the NOL operations were found to have variable hog numbers, 26% were in low intensity agricultural areas, and 54% had increased expansion factors.

FIGURE 2a. CLASSIFICATION OF NOL OUTLIERS BY VARIABILITY OF HOG NUMBERS, AREA STRATIFICATION, AND EXPANSION FACTOR

		Hog Numbers[1]				
		Constant		Variable		
		Expansion Factor[2]		Expansion Factor		
		Same	Increased	Same	Increased	
Area Strata	Strata < 30	33 (26%)	24 (19%)	16 (12%)	22 (17%)	95 (74%)
	Strata 30+	5 (4%)	11 (8%)	5 (4%)	13 (10%)	34 (26%)
		38 (30%)	35 (27%)	21 (16%)	35 (27%)	129 (100%)

[1] An operation was considered variable in hog numbers if the ratio of its highest to lowest number of hogs over the four quarterly surveys was greater than 3. (Operations which had no hogs during any survey are considered variable.)

[2] Expansion factor increases are relative to the June survey expansion factor.

A classification of NOL outliers by three categories found to influence NOL outlier origination. These include Variability in Hog Numbers, Area Stratification and Expansion Factor increases relative to the June survey. Percentages differ slightly from the numbers given in the report since operations which were only sampled in June cannot be categorized with respect to hog number variability.

The second classification was based on an operation's June DE hog total value. The 76 NOL operations fell into one of four possible classifications, conditional on their June DE hog total. Again, the categories emphasize the three potential causes of outlier occurrence given above. Results are shown in Figure 2b below and rows are described relative to types of outlier occurrences.

Row 1 - NOL operations which have no hogs in June and later produce outlier totals for hogs. Often these appear to be seasonal, transient or start-up operations with variable hog numbers throughout the year. Increased expansion factors in follow-on subsampling also influence many of the DE hog totals.

Row 2 - NOL operations which produce hogs in June below the cutoff and later produce outliers in follow-on surveys. Generally, they appear to be operations with small variability in hog numbers across surveys and are often pushed over the cutoff value by the subsampling scheme's increased expansion factor.

FIGURE 2b. CLASSIFICATION OF NOL OUTLIERS CONDITIONAL ON JUNE DE HOG TOTAL

JUNE STATUS OF OPERATION		NUMBER OF OUTLIERS			
		June	Sept	Dec	March
	No Hogs (8 Operations)	0	4	5	4
	DE Total Hogs **less** than outlier cutoff (28 Operations)	0	8	18	12
	DE Total Hogs **greater** than outlier cutoff - operation **resampled** in follow-on survey (29 Operations)	29	21	16	14
	DE Total Hogs expand to **greater** than outlier cutoff - **no resampling** done (11 Operations)	11	-	-	-
	TOTAL NOL OUTLIERS (76 Operations)	40	33	39	30

A categorical distribution of NOL operations producing outliers based on their June DE hog total status. A portion of NOL operations with few or no hogs in June will create outliers in follow-on surveys due to increased expansion values or hog numbers. NOL operations which produce outliers in June consistently continue to produce outliers when resampled in follow-on surveys.

Row 3 - NOL operations which produce outlier values in June. These are operations which are located on light ag-usage land with large associated expansion factors, or are large operations missed by the list-frame, or both.

Row 4 - NOL operations producing outliers in June and not resampled again during the frame-year. These fall into the 40% of the 60/40% split and presumably would behave like Row 3 had they been resampled.

Multiple Outliers From a Single Operation.

A study of both NOL and list operations was conducted to observe the average number of times an operation produced an outlier within a frame-year. More than one outlier from a particular operation during the frame-year was not uncommon. For the five states sampled, a NOL outlier operation produced, on average, 1.89 outlier records over the course of a frame-year. A list outlier operation produced, on average, 1.26 outliers over the course of a frame-year. Thus, NOL operations which produce outlier records seem to have a much greater influence over the frame-year than do list operations which produce outliers. This is especially true since NOL outliers are, on average, larger than list outliers. (This will be shown in the state level results to follow.) These average rates of occurrence include NOL records which were sampled only in June and list records which were rotated out during the frame-year. Thus, the average recurrence rate is probably underestimated or overestimated for a particular state depending on its follow-on sampling scheme and when it entered the 60/40 split program.

Area Adjustment Weights and NOL Outliers.

Individual NOL operations were investigated for any operation which refused or were inaccesibles in June but later produced an outlier record for hogs during the frame-year. However, no NOL operations and/or records of this type were found.

Causes of NOL outlier creation are elusive. Still, pinpointing and eliminating causes for NOL outlier creation is fundamental to maintaining a quality survey series. The subsampling scheme, the variability in hog numbers and the placement of hog production facilities in low intensity agricultural areas all seem to share some responsibility - often with one another. It is important to remember, however, that while many factors aid in the creation of NOL outliers, the first event that must occur is non-overlap with the list frame. Improved list building and maintenance would go a long way in decreasing the impact of NOL outliers. The reproduction of outliers by the same NOL operation within a frame-year shows that any attempt to reduce an operation's influence will affect other potential outliers which might be produced by that operation throughout the frame-year.

Analysis of Variance (ANOVA) Results

State Level ANOVA.

Analysis of variance results for the five selected states produced the findings listed below in Figure 3. A p-value of 0.05 or less is assumed statistically significant.

FIGURE 3. STATE LEVEL ANOVA FOR FRAME-YEAR AND QUARTER EFFECTS

ANOVA MODEL	COLORADO	GEORGIA	IDAHO	ILLINOIS	MICHIGAN
Outlier Total					
Frame-Year	p=0.35	p=0.01*	p=0.01*	p=0.72	p=0.01*
Quarter	p=0.91	p=0.32	p=0.14	p=0.82	p=0.62
Mean Outlier					
Frame-Year	p=0.23	p=0.28	p=0.34	p=0.33	p=0.74
Quarter	p=0.55	p=0.59	p=0.23	p=1.00	p=0.51
Outlier Number					
Frame-Year	p=0.48	p=0.00*	p=0.25	p=0.82	p=0.02*
Quarter	p=0.95	p=0.92	p=0.28	p=0.80	p=0.53

Outlier Total - Outlier Total Per Survey
Mean Outlier - Individual Average Value per Occurrence
Outlier Number - Number of Outliers Occurring per Survey
* Statistically significant at the 0.05 level

Probability values (p-values) for ANOVA test for quarter and frame year. Three states exhibit frame-year effects for Outlier Total per Survey and two states show frame year effects for Outlier Number per Survey. No quarter (seasonal) effects are seen.

A significant frame-year effect for Outlier Total per Survey is present in Georgia, Idaho, and Michigan. This implies that outlier totals within at least one frame-year differ significantly from outlier totals across at least one other frame-year. Outlier Number of Occurrences per Survey also exhibits a significant frame-year effect for Georgia and Michigan. Thus, at least one frame-year for both states produced significantly more or less outliers as compared to the other frame-years. It is unknown whether this difference is due to procedural differences between frame-years (such as sampling, list maintenance or rotation) or whether it is due to changes within the target population. Neither frame-year nor quarter (seasonal) effects are seen in the Mean Outlier per Occurrence. This implies individual outlier magnitude does not vary with time or seasonally. The lack of a quarter

effect in Outlier Totals per Survey and Outlier Number of Occurrences for any of the five states is important. It shows that the RE currently employed, which weights all outliers equally regardless of quarter, is appropriate. Again however, this lack of trend must be monitored since all states are now entered in the 60/40 split program and significant differences in number of outlier occurring or magnitude of outlier may be found in June versus follow-on surveys as more data becomes available. Also, the lack of a quarter effect for all models and states shows that the seasonal trends known to exist in hog production do not affect the number of outliers occurring or their magnitude.

The presence of a frame-year effect for Outlier Total per Survey in three of the states shows the need for an estimator like the RE, which smooths effects of large frame-year outlier totals over time. The presence of a frame-year effect for Outlier Number of Occurrences per Survey implies that the differences in outlier totals by frame-year are due primarily to differences in the number of outliers and not to their size. This suggests one of two approaches. If one believes that an increase or decrease in outlier occurrences indicates hog production trends for that frame-year, the Robust Estimator should take that into consideration within its second component at the state level. One possibility would be to take the product of an average outlier magnitude computed for that state and the number of outliers occurring (possibly by type). Alternatively, if one believes that increases in outlier numbers are the result of the sampling scheme, list quality or other transient phenomenon, then one should allow the RE to estimate the outlier component using all outlier data across several years. Additional study is required to find which approach is most appropriate.

Pseudo-Regional Five-State and National Level ANOVA.

The outlier data for the five states were combined to form a pseudo-regional five-state level. At this level, similarities can be seen between the five-state combined and the national summary data in the outcome of ANOVA tests. Tests were performed for frame-year and quarter effect on outlier totals and number of outliers occurring per survey.

FIGURE 4. COMPARISON OF COMBINED 5-STATE AND NATIONAL 48-STATE ANOVAs FOR OUTLIER TOTAL AND NUMBER OF OCCURRENCE

ANOVA MODEL	5-STATE SAMPLE	NATIONAL 48-STATE
Outlier Total		
Frame Year	p=0.03*	p=0.97
Quarter	p=0.40	p=0.58
Surveys Used	n=13	n=14
Outlier Number		
Frame Year	p=0.20	p=0.27
Quarter	p=0.86	p=0.23
Surveys Used	n=13	n=13

* Statistically significant at the 0.05 level

The combined 5-state ANOVA finds a frame-year effect for outlier total not seen in the national summary data. All other tests show no effects for either data set.

Figure 4 below shows Outlier Totals at the 5-state level have a significant trend in frame-year while none is found at the 48-state national level. For Outlier Number, neither the 5-state nor 48-state ANOVA tests detected a trend in frame-year or quarter. Again, this is good news for the Robust Estimator, since it appears that the number of outliers is not changing over time - either across years or seasonally. The presence of a frame-year effect

for Outlier Total at the 5-state level again implies the need for the RE which smooths large outliers occurring in a frame-year across several frame-years.

At the national level outliers appear to be both equal and consistent for survey totals and numbers, both across frame-years and quarters. This leads one to believe that the Robust Estimator, which uses all outlier information available, will produce a better national hog total estimate than the DE estimator. The use of outlier information gathered over several previous surveys should help the RE produce an estimate which is not only accurate but precise. Where such information is consistent from survey to survey (such as at the national level) RE estimates will be comparable with the direct expansion estimate but

variance of the estimate series should be reduced (ie., the RE should be as accurate as the survey direct expansion estimate and more precise). Where such information is not consistent (such as at the state level) the RE can provide not only a more accurate estimate, but also a more precise estimate than the survey direct expansion estimate.

Comparison of Two State Level Outlier Distributions

All outliers, both list and NOL, were grouped by state to quantify state level distribution characteristics. Since each state is unique and has its own hog production properties, it is not surprising to find many differences in the five selected states' outlier distributions. Because of these differences, it is often difficult to draw comparisons across states. Summary statistics by type of record (list or NOL) for each of the five states were computed for 1) the total number of outliers occurring over the four frame-years, 2) the average of number of outliers occurring per survey, 3) the average outlier DE hog total per occurrence and 4) the average outlier DE hog total per survey. The results for all five states can be found in Appendix E. A comparison of the two most comparable states in total hog production, Georgia and Michigan, will be shown here to exemplify the influence that the cutoff value has on the outlier distribution at the state level, and the differences that can occur between states. Figure 5 below shows the results.

For NOL records, Georgia and Michigan appear to have similar numbers of outliers occurring per survey, though average magnitudes differ. This is largely due to the three 400,000-plus expanded hog records Georgia had during the 1989 frame-year.

It is primarily within the list records that noticeable differences prevail between the two states. List outlier records occur at a much greater rate for Michigan than Georgia. This difference occurs though Georgia and Michigan produce comparable survey DE hog totals (an average of 1.24 versus 1.23 million head per survey over the 15 quarters). This discrepancy is largely a function of the outlier cutoff value for each state (though differences in operation characteristics between the two states may have some effect).

Michigan's outlier distribution is composed of a much greater percentage of list records than Georgia's because its cutoff value cuts farther into its list record distribution. The extreme right tail of the outlier distribution for both states is made up entirely of NOL records due to their large expansion values as compared to list records. When the cutoff value decreases toward the individual DE hog total record average, the outlier distribution changes from one composed predominantly of NOL records to one composed predominantly of list records, as in Michigan's case. List records will be included faster than NOL records because of their predominance in the survey and because they tend to be grouped tighter about the overall record average due to their lower expansion factors. A graphical

FIGURE 5. COMPARISON OF GEORGIA AND MICHIGAN OUTLIER DISTRIBUTIONS COMPILED FROM 1987-1990 DATA

	GEORGIA (15 Surveys)	MICHIGAN (14 Surveys)
OUTLIER CUTOFF VALUE	(25,000)	(15,000)
NOL		
NUMBER OF OUTLIERS (All Surveys)	36	31
AVERAGE OUTLIER -		
Occurring per Survey	2.40	2.21
Size per Occurrence	108,214	30,035
Total per Survey	259,715	66,505
LIST		
NUMBER OF OUTLIERS (All Surveys)	3	87
AVERAGE OUTLIER -		
Occurring per Survey	0.20	6.21
Size per Occurrence	27,669	21,782
Total per Survey	5,533	135,361
COMBINED NOL/LIST		
NUMBER OF OUTLIERS (All Surveys)	39	118
AVERAGE OUTLIER -		
Occurring per Survey	2.60	8.41
Size per Occurrence	102,019	23,950
Total per Survey	265,248	201,866
NOL % of Outlier Total	92%	26%
LIST % of Outlier Total	8%	74%
% of Survey Total	21.4%	16.0%
NOL % of Survey Total	21.0%	5.0%
LIST % of Survey Total	0.4%	11.0%

Georgia's and Michigan's outlier distributions compiled over four frame-years show similar combined NOL/list outlier totals and percentages of survey totals. However, within each state's combined outlier distribution the proportion of list and NOL records and the average size of those records is very different.

comparison of Georgia and Michigan outlier distributions at current cutoff values (Figures 6a and 6b) and with an equivalent cutoff value of 25,000 (Figure 6c), are shown below. Note that the vertical frequency axes for current state distributions are not of the same scale. Once cutoffs are equated, Michigan continues to have proportionally more list records but nowhere near its previous amount. Distributions are now of the same scale.

NOL records for the two states occur at similar rates but have different magnitudes (See Figure 5). This difference is probably related to Georgia's large outliers of 1989. Alternatively, the two states' list records occur at different rates with similar magnitudes. This is due primarily to the state cutoff value. Both these circumstances occur though the states have comparable total hog production. These differences combine to produce very distinct compositions for each state's outlier distribution. At the national level it is possible that more equitable distribution compositions could be formed for all states by examining alternative cutoff values. If total hog production is used to define outliers, then one would expect states with similar production total characteristics to produce comparable numbers and types of outliers over time. Better equating of outlier distributions could increase the RE's effectiveness.

FIGURE 6a. OUTLIER DISTRIBUTION FOR EXPANDED HOGS BY TYPE
STATE = MICHIGAN - CUTOFF=15000

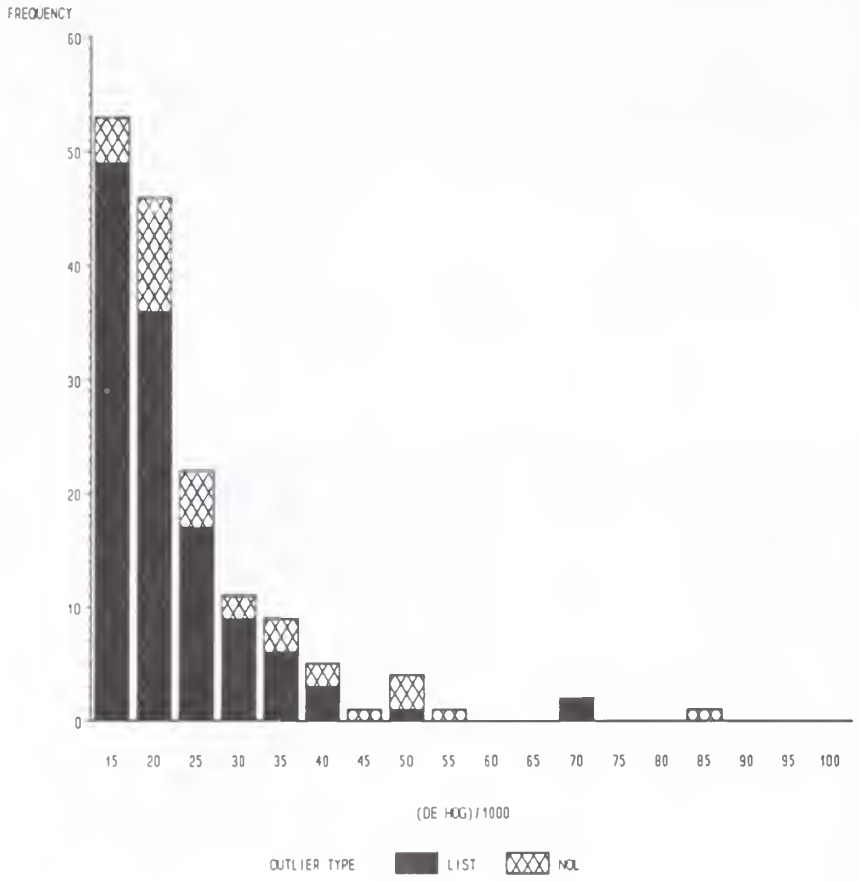
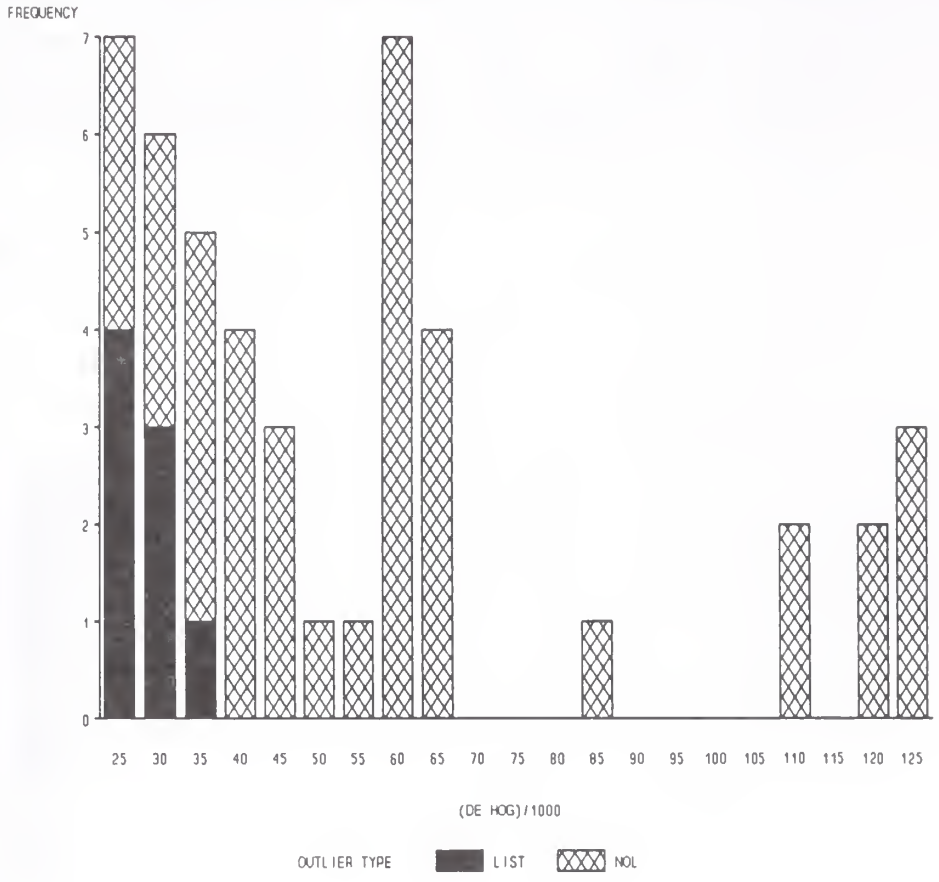
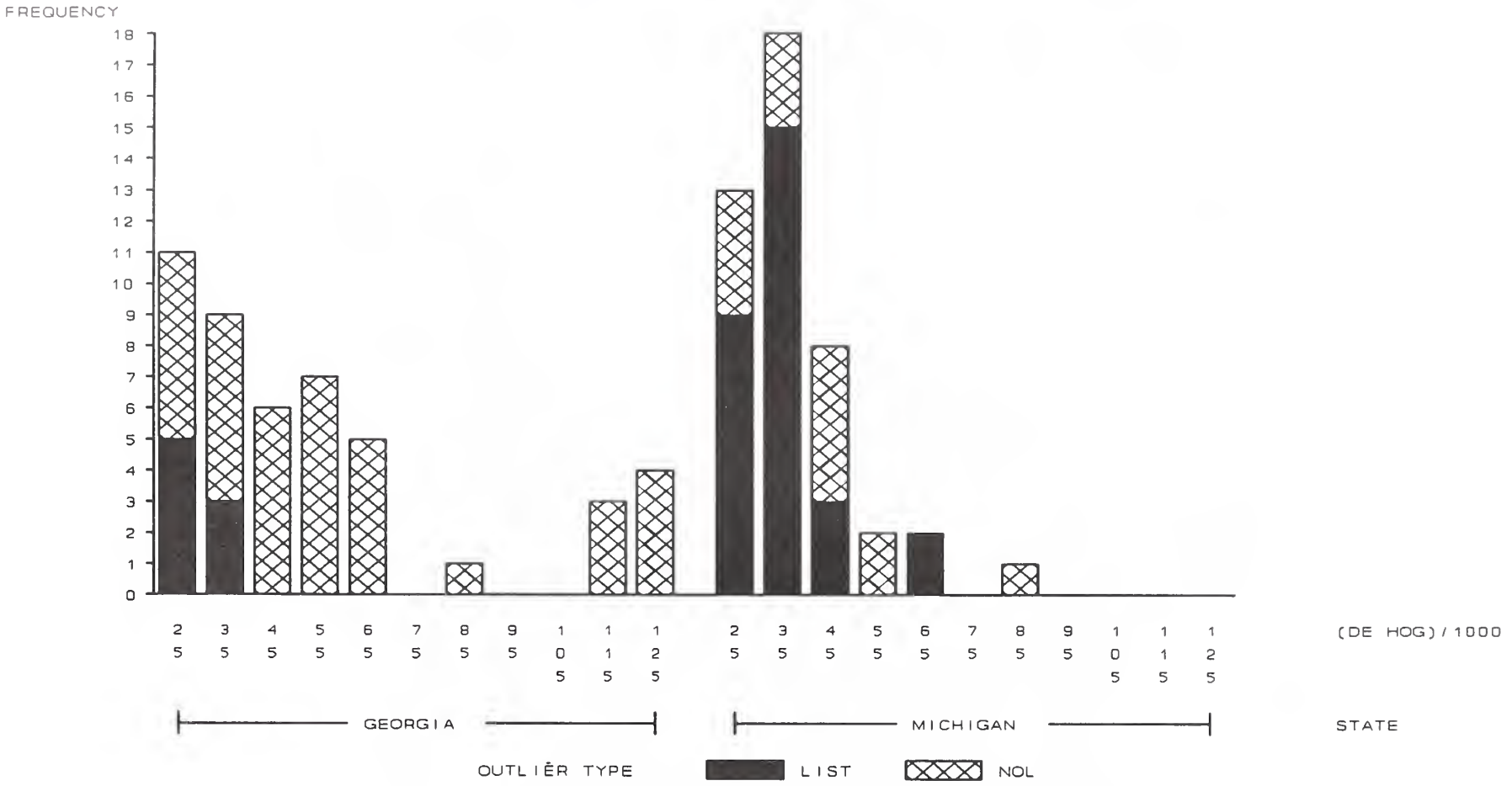


FIGURE 6b. OUTLIER DISTRIBUTION FOR EXPANDED HOGS BY TYPE
STATE = GEORGIA - CUTOFF=25000



A comparison of Michigan and Georgia outlier distributions for all outliers occurring from June 1987 through December 1990. Figures 6a and 6b use current cutoff values for each state to define outliers.

FIGURE 6c. COMPARISON OF OUTLIER DISTRIBUTIONS FOR AN EQUAL CUTOFF VALUE OF 25000



A comparison of Michigan and Georgia outlier distributions for all outlier occurring from June 1987 through December 1990. Figure 6c uses an equated cutoff value of 25,000 to define outliers for both states.

CONCLUSIONS

Many of the origins of individual NOL hog outliers are hidden. Indeed, there may exist as many causes as there are outliers themselves. Still, attempts must be made to group like causes and reduce outlier numbers to maintain the quality of the survey series. Three potential causes of NOL outliers were found in the data analyzed in this report. They include: the follow-on sampling scheme, the transitory nature of hog production leading to poor follow-on stratification, and variability in hog operation locations with respect to the area-frame stratification procedure.

Over one-fourth of all NOL operations which produced an outlier were found to do so as a direct result of increased expansion factors with little change in hog production characteristics. One-fourth of the NOL operations producing outliers lacked hogs in at least one survey, yet produced enough hogs in a later survey during the same frame-year to create an outlier. Over forty percent had more than three times as many hogs during one survey, as compared to the other surveys during the frame-year. Finally, nearly thirty percent of NOL operations producing outliers were found on stratum 30 or greater.

Though there are no simple answers to the reduction in outliers produced by the three causes listed above, perhaps the following suggestions could be studied further.

- Follow-on sampling schemes for the NOL population should include as much of the original June base survey NOL sample as feasibly possible. Generally, respondent burden and data collection costs become the deciding factor in sample size and allocation for follow-on surveys. Use of either a peak hog inventory question for area respondents in June or use of the most recent survey information regarding hogs, or both, might prove effective in reducing outliers. This would be could potentially lead to slight increases in sample size and respondent burden.
- Any improvement in area stratification or area sample allocation with respect to livestock would help to reduce the number of hog outliers produced. It is difficult to conceive of another area stratification method other than land usage for the area-frame. Perhaps additional stratification, if found, or larger allocations of a state's total sample to higher area strata in large hog producing states would provide improvement. Again, sampling efficiency with respect to the overall survey becomes the deciding factor.
- Lastly, increased list building and maintenance would reduce the number of NOL operations and their associated large expansion factors. In fact, reduction of NOL tracts in June would result in a multiple reduction in outliers since individual NOL operations very often produce more than one outlier in a frame-year. Presently, a major list building initiative is underway within NASS which will provide an indication of coverage improvements versus costs.

The presence of a frame-year effect at the state level for outlier total shows the need for the Robust Estimator. The presence of a frame-year effect at the state level for outlier number of occurrences implies a modified RE which accounts for the number of outlier occurring could provide more efficient estimates when this effect is due to production changes. If the differences in the number of outlier occurring is due to survey design variability the current RE should prove an adequate estimator. At the national level the survey outlier totals and survey outlier numbers occurring appear to be steady across frame-years and quarters. The lack of any frame-year or quarter effect shows that all surveys ought to be used to calculate the outlier component, as is presently done, and that

the Robust Estimator should continue to be used without modification. The lack of any quarter effect at the state and national level shows that seasonal hog production characteristics are not being felt by the outliers themselves.

These results are based on analyses of NOL outliers from June 1987 through December 1990. A new sampling scheme has been implemented by Sample Design Section for the 1991 frame year which uses the same NOL sample for all follow-on surveys within the frame-year. The purpose of this sampling scheme is to produce more stable NOL indications across the frame-year. The impact on NOL outlier characteristics is expected to be minimal, but actual results and their effect on the Robust Estimator should be assessed. What this analysis does not answer is the question of differences in state outlier cutoffs. How much do these cutoff values affect the outlier component of the Robust Estimator at the different levels? Can outlier distributions be equated across states relative to hog production? Would it improve the Robust Estimator at the regional or national level if it were done? It is apparent that there are differences in the cutoff values, and the distributions they define, at the state level. Further investigation into outlier cutoff values and the optimum number of surveys to use in calculating the second component of the RE is required to answer these questions and produce the best estimator possible. The analysis shows that the Robust Estimator is a needed tool that possess the ability to produce reliable estimates at any level.

RECOMMENDATIONS

Based on the results of this study, it is recommended that:

1. At the national level, up to 15 quarters of outlier data (the most this report can justify) should be used to compute the second component of the Robust Estimator.
2. Investigation continue at the state level comparing the simple Robust Estimator with a modified Robust Estimator which recognizes the frame-year effect for outlier number of occurrences and their effectiveness at setting state level estimates.
3. Sample design be examined to seek ways to reduce the number of NOL outliers by reducing follow-on expansion factors or the number of NOL records. Follow-on NOL expansion factor reduction might be accomplished by stratifying on the peak number of hogs an operation expects over the course of the frame-year, asked during the June survey, or by using the most recent previous survey data available, and to stratify an operation in the current survey. The number of NOL records could be reduced through increased list building and maintenance.
4. Investigation be instituted to study the impact that different state cutoff values and number of surveys used have on improving the Robust Estimator. Changes in cutoff values could result in an increase in data processing if both old and new cutoff values are maintained.

REFERENCES

Keough, Gary & Perry, Charles R.; A Review and Evaluation of Unusual Outlier Component Detection and Robust Estimation Procedures Used by the December 1989 Hog Board. NASS Staff Report No. SRB-91-05; United States Department of Agriculture: Washington, D.C.; April 1991.

Nealon, John Patrick; Review of the Multiple and Area Frame Estimators. NASS Staff Report SF/SRB No. 80; United States Department of Agriculture: Washington, D.C.; March 1984.

Pafford, Brad, et al.; The Concept of Accuracy and How to Approach Reducing the Total Survey Error of NASS Survey Statistics, with an Example from the Agricultural Survey Program. NASS Survey Quality Team Report No. 90-1; United States Department of Agriculture: Washington, D.C.; October 1990.

Rumburg, Scot; Retrieval of Previous Survey Data Sets for the Hog Outlier Analysis Report: A Narrative Approach to the Martin Marietta Data System and General Edit Procedure. Estimates Research Section. April 1991.

Thomas, David R., Perry, Charles R. & Viroonsri, Boonchai; Estimation of Totals for Skewed Populations in Repeated Agricultural Surveys: Hogs and Pigs. NASS Staff Report No. SRB-90-02, United States Department of Agriculture: Washington, D.C.; April 1990.

APPENDIX A

NOL FOLLOW-ON SUBSAMPLING SCHEME (Valid through March 1991)

Both September and December NOL subsampling schemes are based on **reported June data.** The March NOL sample is a subsample of the December NOL sample. These sampling schemes apply only to the **60% segments available for Ag Surveys.**

SEPTEMBER NOL SUBSAMPLING SCHEME

Summary Stratum	Summary Stratum Description	Approx. Sampling Proportion
1	NOL, Positive hogs, stocks, small grains (includes missing or incomplete data), or hog intentions	1
2	NOL, None of the above, but positive or missing capacity	1/2
3	NOL, None of the above ("zero")	1/5
4	Overlap and Non-ag	0

DECEMBER NOL SUBSAMPLING SCHEME

Summary Stratum	Summary Stratum Description	Approx. Sampling[1] Proportion
1	NOL, "Large" (the upper 90 percentile) of expanded hogs, capacity, cropland or chickens	1
2	NOL, "Medium" (greater than median, but less than the 90 percentile) of expanded hogs	1/2
3	NOL, "Medium" expanded capacity, cropland	1/2
4	NOL "Small" (less than median) expanded hogs or intentions	1/2
5	NOL "Small" expanded capacity, cropland or chickens, or missing or incomplete data	1/2
6	NOL, none of the above ("zero") or Non-ag with potential	1/5
7	Overlap and Non-ag with no potential	0

[1] Sampling proportions vary depending upon the tract's cross-classify status with January Cattle.

MARCH NOL SUBSAMPLING SCHEME

Summary Stratum	Summary Stratum Description	Approx. Sampling Proportion
1	Positive stocks in December (includes missing or incomplete data)	1
2	None of the above, but positive or incomplete for hogs in December	1
3	None of the above, but positive or incomplete for selected crops in December	1
4	None of the above ("zero" in December)	1/2

APPENDIX B

RANK OF TOTAL HOG INVENTORIES AND ASSOCIATED CUTOFF VALUE BY STATE

STATE	RANK	OUTLR CUTOFF	TOTAL 1990 HOG & PIG INVENTORY (Head / 1000)	STATE	RANK	OUTLR CUTOFF	TOTAL 1990 HOG & PIG INVENTORY (Head / 1000)
IA	1	80000	14000	CA	26	5000	180
IL	2	50000	5700	MD	27	5000	162
MN	3	40000	4450	MS	28	10000	149
NE	4.5	40000	4300	FL	29	10000	130
IN	4.5	40000	4300	AZ	30	5000	110
MO	6.5	40000	2800	NY	31	5000	103
NC	6.5	30000	2800	OR	32	5000	80
OH	8	25000	2000	ID	33	5000	60
SD	9	25000	1770	WA	34	5000	56
KS	10	20000	1500	LA	35	5000	50
MI	11	15000	1250	HI	36	3000	36
WI	12	25000	1150	UT	37.5	5000	33
GA	13	25000	1100	MA	37.5	5000	33
PA	14.5	20000	920	DE	39	5000	31
KY	14.5	20000	920	WV	40	5000	30
AR	16	10000	760	NM	41	5000	27
TN	17	20000	620	NJ	42	5000	25
TX	18	20000	550	WY	43	5000	20
VA	19	15000	430	NV	44	4000	14
SC	20	10000	410	ME	45	3000	9.9
AL	21	10000	400	NH	46	3000	9.3
CO	22	10000	300	CT	47	5000	6.9
ND	23	5000	265	RI	48	3000	5.3
OK	24	5000	215	VT	49	3000	5.0
MT	25	5000	185	AK[1]	50	-	1.2

States underlined represent ones selected for this study

[1] No outlier cutoff value has been determined for Alaska

APPENDIX C

Data Acquisition, Reproduction and Summarization

Data Acquisition.

Only cumulative statistics are maintained on expanded hog totals and, in particular, outlier expanded hog totals. These statistics are available within the Estimates Division/Multiple Frame Survey Branch (ED/MFSB) of NASS. They include sum totals of all usable outlier records by state and survey, and the number of usable records found to exceed the cutoff, again by state and survey. Likewise, state and national totals for all usable records are computed, and list-frame and non-overlap (NOL) sample sizes.

Since the variable of interest, expanded hog total (DE hog total), is a function of several survey variables and not available directly from the raw data, it must be computed after retrieving the raw data. The primary source of previous quarterly agricultural survey data is the Martin Marietta Data System (MMDS) where past survey data have been archived. Currently work is underway at NASS/SRB/Technology Resources Section to archive and maintain survey data on-site. This should hasten future data retrieval.

Data Reproduction.

The raw data obtained from MMDS must first be edited prior to analysis. This generally involves two steps - combining of record information to the tract level and determination of usable records. Once the raw data undergoes the two step-edit procedure, simple statistics can be produced which should agree with the known cumulative summary statistics maintained by ED/MFSB. The equating of these statistics ensures that the data to be analyzed are a reasonable facsimile of the data used to produce the summary statistics.

Estimates Division uses a general edit program to "clean" survey data. It combines all records to the tract level and obtains all usable records. Because of many subtle differences in the actual summary edit versus hand editing procedures, a data set produced outside the general edit program will not, without excessive time and effort, exactly match the data set used by Estimates Division. However, the data sets which were produced represent near-exact replicas for outlier data and a reasonable representation for the total data itself. Statistical comparisons of the data sets created for analysis versus Estimates Division's summary statistics are shown below. These comparisons include DE hog totals for all data, DE hog totals for outliers, sample sizes by strata, numbers of usables, and number of NOL records. Comparisons of actual versus reproduced 5-state summary data for June 1987 to December 1990 are shown below.

REPRODUCED 5-STATE SUMMARY DATA VERSUS ACTUAL (OBSERVED) SUMMARY DATA (1987-88) C-2

YR/QTR	ST	SUMMARY EXPANDED HOG TOTAL	OBSERVED EXPANDED HOG TOTAL	DIFFERENCE [OBS-SMMRY] (HOG TOTAL)	SUMMARY EXPANDED OTLR TOTAL	OBSERVED EXPANDED OTLR TOTAL	DIFFERENCE [OBS-SMMRY] (OTLR TOTAL)	# SMMRY OTLR	# OBS OTLR	DIFF #[O-S]	SMMRY OTLR % OF SMMRY HOG TOTAL
1987 FRAME-YEAR											
87/Q2	CO	186435	186436	1	0	0	0	0	0	0	0.00%
(JUN)	GA	1016571	1016568	-3	0	0	0	0	0	0	0.00
	ID	51815	51815	0	0	0	0	0	0	0	0.00
	IL	4943599	4943764	165	134054	134054	0	2	2	0	2.71
	MI	1029308	1029308	9	139610	139610	0	6	6	0	13.56
87/Q3	GA	1171018	1170656	-362	111540	111540	0	1	1	0	9.53%
(SEP)	IL	5321824	5321048	-776	461942	461942	0	5	5	0	8.68
87/Q4	CO	264712	263101	-1611	61027	61028	1	3	3	0	23.05%
(DEC)	GA	1176984	1176386	-598	144096	144096	0	2	2	0	12.24
	ID	72083	72083	0	7310	7310	0	1	1	0	10.14
	IL	5153444	5153412	-32	476979	476979	0	4	4	0	9.26
	MI	1300727	1300589	-138	141164	141164	0	6	6	0	10.85
88/Q1	CO	221098	221051	-47	42075	42076	1	3	3	0	19.03%
(MAR)	GA	1090491	1090235	-256	58373	58373	0	1	1	0	5.35
	ID[1]	80009	*	*	18552	*	*	3	*	*	23.19
	IL	5491435	5491437	2	687064	687065	1	6	6	0	12.51
	MI	1069442	1069289	-153	159553	159553	0	6	6	0	14.92
1988 FRAME-YEAR											
88/Q2	CO	205065	205066	1	10024	10024	0	1	1	0	4.89%
(JUN)	GA	1160198	1182246	22048	96713	96713	0	3	3	0	8.34
	ID	95507	95507	0	28091	28091	0	3	3	0	29.41
	IL	5896453	5896453	0	375726	375726	0	5	5	0	6.37
	MI	1322958	1322958	0	222193	222193	0	10	10	0	16.80
88/Q3	CO	224570	224568	-2	10904	10904	0	1	1	0	4.86%
(SEP)	GA	1297412	1297410	-2	113023	113022	-1	2	2	0	8.71
	ID	111655	111650	-5	44241	44237	-4	2	2	0	39.63
	IL	5893801	5893801	0	362292	362292	0	5	5	0	6.15
	MI	1434974	1446974	12000	216802	216802	0	8	8	0	14.98
88/Q4	CO	219703	219703	0	10502	10502	0	1	1	0	4.78%
(DEC)	GA	1158196	1156083	-2113	104035	104034	-1	2	2	0	8.98
	ID	78971	78971	0	0	0	0	0	0	0	0.00
	IL	5346325	5345372	-953	256587	256587	0	4	4	0	4.80
	MI	1213377	1213377	0	74767	74767	0	3	3	0	6.16
89/Q1	CO	229579	229579	0	0	0	0	0	0	0	0.00%
(MAR)	GA	1117552	1117550	-2	130486	130485	-1	2	2	0	11.68
	ID	69061	69061	0	8190	8190	0	1	1	0	11.86
	IL	4951179	4947059	-4120	288471	288472	1	4	4	0	5.83
	MI	1140694	1140695	1	190076	190076	0	6	6	0	16.67

[1] Observed Data Set Not Retrievable

REPRODUCED 5-STATE SUMMARY DATA VERSUS ACTUAL (OBSERVED) SUMMARY DATA (1988-89) C-3

YR/QTR	ST	SUMMARY EXPANDED HOG TOTAL	OBSERVED EXPANDED HOG TOTAL	DIFFERENCE [OBS-SMMRY] (HOG TOTAL)	SUMMARY EXPANDED OTLR TOTAL	OBSERVED EXPANDED OTLR TOTAL	DIFFERENCE [OBS-SMMRY] (OTLR TOTAL)	# SMMRY OTLR	# OBS OTLR	DIFF #[O-S]	SMMRY OTLR % OF SMMRY HOG TOTAL
1989 FRAME-YEAR											
89/Q2	CO	257730	257730	0	47462	47462	0	3	3	0	18.42%
(JUN)	GA	1237391	1237391	0	245056	245056	0	4	4	0	19.80
	ID	103419	103419	0	40201	40201	0	4	4	0	38.87
	IL	6039423	6039423	0	529265	529265	0	5	5	0	8.76
	MI	1273731	1273731	0	199359	199359	0	11	11	0	15.65
89/Q3	CO	240357	240357	0	28254	28254	0	2	2	0	11.76%
(SEP)	GA	1482429	1482423	-6	598574	598569	-5	4	4	0	40.38
	ID	123104	123105	1	71356	71356	0	4	4	0	57.96
	IL	5951168	5954542	3374	508203	508203	0	7	7	0	8.54
	MI	1262279	1262279	0	159390	159390	0	7	7	0	12.63
89/Q4	CO	227025	226996	-29	0	0	0	0	0	0	0.00%
(DEC)	GA	1669773	1669765	-8	859081	859073	-8	4	4	0	51.45
	ID	93150	93293	143	42096	42096	0	1	1	0	45.19
	IL	5306544	5309563	3019	254649	254649	0	4	4	0	4.80
	MI	1247205	1250205	3000	201282	201866	584	7	7	0	16.14
90/Q1	CO	210978	211460	482	0	0	0	0	0	0	0.00%
(MAR)	GA	2001956	2001945	-11	1159054	1159054	0	4	4	0	57.90
	ID	106141	106151	10	50593	50593	0	2	2	0	47.66
	IL	4806941	4797672	-9269	317650	317650	0	5	5	0	6.61
	MI	1132595	1132678	83	212746	212746	0	9	9	0	18.78
1990 FRAME-YEAR											
90/Q2	CO	270996	270996	0	0	0	0	0	0	0	0.00%
(JUN)	GA	1097802	1097302	-500	127952	127952	0	4	4	0	11.66
	ID	75584	75584	0	13330	13330	0	1	1	0	17.64
	IL	5262127	5376057	61	292183	292182	-2	4	3	-1	5.44
	MI	1348932	1348932	0	292403	292403	0	12	12	0	21.68
90/Q3	CO	285147	285146	-1	0	0	0	0	0	0	0.00%
(SEP)	GA	1062059	1062058	-1	120327	120326	-1	3	3	0	11.33
	ID	59492	59492	0	6942	6942	0	1	1	0	11.67
	IL	5648985	5638245	-1074	322817	322818	1	3	3	0	5.72
	MI	1303431	1303432	1	311917	311917	0	13	13	0	23.93
90/Q4	CO	298898	298779	-119	10487	10486	-1	1	1	0	3.51%
(DEC)	GA	921006	921005	-1	110439	110438	-1	3	3	0	11.99
	ID	44305	44305	0	0	0	0	0	0	0	0.00
	IL	5652467	5652565	98	908421	908419	-2	9	9	0	16.07
	MI	1207167	1207167	0	304565	304565	0	14	14	0	25.23

APPENDIX D

Description of Analysis of Variance (ANOVA)

Analysis of Variance (ANOVA) checks to see if a specific model fits the data better than a model where each value is estimated to be the overall mean. This is done by looking at the reduction in variance for a model which treats each subgroup as a unique population versus a model which treats all data as belonging to the same population. Since any model with an added parameter will nearly always explain more of the variance within the data (and will always explain as much as the more general model) the real test is whether the added parameter explains enough of the variance to justify it statistically.

The null hypothesis model versus the alternative hypothesis model for a single effect is:

$$H_o\colon\ y_i = \beta_0 + \epsilon_i \quad \textit{versus} \quad H_a\colon\ y_{ij} = \beta'_{0j} x_{ij} + \epsilon_{ij}\ .$$

H_o is a regression model with slope zero and y intercept equal to the average for all data (β_o). The model's ability to fit the data is given by the sum of the squared residual errors (SSE) for H_o ($\Sigma\epsilon_i^2$). Alternatively, H_a is a series of regressions on the j subgroups of interest (say, all data which occurs in June, December, September or March in a seasonality test) where x_{ij} is an indicator variable equal to one if the i^{th} data value is in the j^{th} subset and is otherwise equal to zero. This creates a series of j regression lines which have zero slope and y intercept equal to the subgroup mean (β'_{oj}) (i.e., June, December, September and March averages). This minimizes the residual SSE for H_a ($\Sigma\epsilon_{ij}^2$) of the data for this effect (the total of the subgroup SSEs). Frame-year effect is handled in the same manner with subgroups formed from frame 1987, frame 1988, etc.

The ratio of the residual SSE of H_o is then compared to the residual SSE of H_a to test for statistical significance of the added effect. This result is conditional on the number of observations in each group and the number of groups. If any subgroup is significantly different from the overall mean the test will reject H_o since a significant amount of variance can be explained by the added effect. Subgroups can then be compared pair-wise (say, June and December) to see which subgroups differed significantly.

ANOVA maintains an overall (experiment-wise) significance level for all subgroup comparisons and allows for unequal sample sizes within subgroups (unbalanced design). Thus an alpha level which accepts an effect as present (rejects H_o) is equal to or less than the actual alpha level given. The outcome for an ANOVA test is usually denoted by a p-value. This value represents the probability that a tested effect is present - the conditional probability of observing a distribution value as large or larger, given H_o is true versus what was found given the ratio of SSE H_o and SSE H_a. The p-value is the value at which the experiment would reject H_o and a small p-value implies that H_o is false and the effect is present. It is also a measure of the amount of risk you are assuming for accepting H_a. With a p-value of .05 for example, on average the test will detect 5 differences out of 100 where, in reality, there is no difference. For purposes of this report a p-value of .05 or less is considered statistically significant. In general, a significance level for p-values also could be set lower or higher than .05 depending on the desire to detect differences between populations and the risk one is willing to accept.

APPENDIX E

INDIVIDUAL NOL OPERATION RESPONSES
QUARTERLY AG SURVEYS FROM JUNE 1987 THROUGH MARCH 1988
NON-OVERLAP OBSERVATIONS ONLY

STATE	STRATA	IDENT	SURVEY	MPRESHOG	MRESPHOG	MEXPFCTR	LHOGTOTL	EXPDHGTL
COLORADO (CUTOFF=10000)								
8	13	8-87-01A	Q287	0	1	96.40	272	9757
8	13	8-87-01C	Q487	4	2	160.67	418	24990
8	13	8-87-01D	Q188	4	2	160.67	328	19609
8	**34**	**8-87-02A**	**Q287**	**0**	**4**	**53.17**	**10**	**532**
8	**34**	**8-87-02C**	**Q487**	**4**	**2**	**212.67**	**12**	**2552**
8	**34**	**8-87-02D**	**Q188**	**4**	**2**	**212.67**	**57**	**12122**
8	47	8-87-03A	Q287	0	1	96.60	72	3478
8	47	8-87-03C	Q487	4	2	322.00	138	22218
8	47	8-87-03D	Q188	4	4	322.00	0	0
GEORGIA (CUTOFF=25000)								
13	13	13-87-04A	Q287	0	1	146.40	528	21505
13	13	13-87-04B	Q387	4	2	146.40	458	18654
13	13	13-87-04C	Q487	4	3	146.40	635	25863
13	13	13-87-04D	Q188	4	3	146.40	0	0
13	**40**	**13-87-05A**	**Q287**	**0**	**2**	**185.90**	**0**	**0**
13	**40**	**13-87-05B**	**Q387**	**4**	**2**	**185.90**	**600**	**111540**
13	**40**	**13-87-05C**	**Q487**	**4**	**2**	**185.90**	**636**	**118232**
13	**40**	**13-87-05D**	**Q188**	**4**	**2**	**185.90**	**314**	**58373**

(For ease of identification, unique operations appear alternately in bold and non-bold type face.

DEFINITIONS FOR VARIABLES

VARIABLE	DEFINITION
STATE	State FIPS Code
STRATA	Area Stratum of the Operation
IDENT	Record ID (State - Frame-year - NOL Record #)
SURVEY	Survey Code - Quarter and Year (Q2=June, Q3=Sept, Q4=Dec, Q1=Mar), second two digits are year of survey
MPRESHOG	Hog Presence Code for June Survey
MRESPHOG	Survey Response Code for June Survey
MEXPFCTR	Area Expansion Factor
LHOGTOTL	Hog Total for the Operation
EXPDHGTL	DE Hog Total for the Operation

NOTE: Definitions for June code variables MPRESHOG, MRESPHOG, and STRATA are given at the end of Appendix E.

INDIVIDUAL NOL OPERATION RESPONSES FOR
QUARTERLY AG SURVEYS FROM JUNE 1987 THROUGH MARCH 1988
NON-OVERLAP OBSERVATIONS ONLY

STATE	STRATA	IDENT	SURVEY	MPRESHOG	MRESPHOG	MEXPFCTR	LHOGTOTL	EXPDHGTL
IDAHO (CUTOFF=5000)[1]								
16	15	16-87-06A	Q287	0	4	93.60	0	0
16	15	16-87-06C	Q487	4	2	304.57	24	7310
ILLINOIS (CUTOFF=50000)								
17	11	17-87-07A	Q287	1	6	173.00	81	6715
17	11	17-87-07B	Q387	4	3	173.00	1835	152114
17	11	17-87-07C	Q487	4	3	346.00	1845	305886
17	11	17-87-07D	Q188	4	3	346.00	1981	328433
17	**11**	**17-87-08A**	**Q287**	**0**	**4**	**173.00**	**407**	**30011**
17	**11**	**17-87-08B**	**Q387**	**4**	**2**	**173.00**	**0**	**0**
17	**11**	**17-87-08C**	**Q487**	**4**	**2**	**173.00**	**745**	**54935**
17	**11**	**17-87-08D**	**Q188**	**4**	**3**	**173.00**	**442**	**32592**
17	11	17-87-09A	Q287	0	1	173.00	97	16781
17	11	17-87-09B	Q387	4	2	173.00	164	28372
17	11	17-87-09C	Q487	4	2	321.29	124	39840
17	11	17-87-09D	Q188	4	2	321.29	190	61045
17	**12**	**17-87-10A**	**Q287**	**0**	**1**	**200.60**	**946**	**68069**
17	**12**	**17-87-10B**	**Q387**	**4**	**2**	**200.60**	**794**	**57132**
17	**12**	**17-87-10C**	**Q487**	**4**	**3**	**200.60**	**794**	**57132**
17	**12**	**17-87-10D**	**Q188**	**4**	**3**	**200.60**	**794**	**57132**
MICHIGAN (CUTOFF=15000)								
26	12	26-87-11A	Q287	0	1	191.40	134	15625
26	12	26-87-11C	Q487	4	2	191.40	113	13176
26	12	26-87-11D	Q188	4	2	191.40	44	5131
26	**20**	**26-87-12A**	**Q287**	**0**	**1**	**220.80**	**10**	**2208**
26	**20**	**26-87-12C**	**Q487**	**4**	**1**	**441.60**	**17**	**7507**
26	**20**	**26-87-12D**	**Q188**	**4**	**1**	**441.60**	**38**	**16781**

QUARTERLY AG SURVEYS FROM JUNE 1988 THROUGH MARCH 1989
NON-OVERLAP OBSERVATIONS ONLY[2]

COLORADO (CUTOFF=10000)								
8	13	8-88-13A	Q288	0	1	96.40	161	10024
GEORGIA (CUTOFF=25000)								
13	13	13-88-14A	Q288	0	1	146.40	998	35407
13	13	13-88-14B	Q388	3	2		.	.
13	13	13-88-14C	Q488	3	6		.	.
13	13	13-88-14D	Q189	3	6		.	.

[1] ONLY 2 SURVEYS (JUNE AND DECEMBER 1987) ARE AVAILABLE FOR IDAHO IN THE 1987 FRAME-YEAR.

[2] CODE DEFINITIONS FOR THE 1988, 1989 AND 1990 FRAMES ARE LOCATED AT THE END OF APPENDIX E.

INDIVIDUAL NOL OPERATION RESPONSES FOR
QUARTERLY AG SURVEYS FROM JUNE 1988 THROUGH MARCH 1989
NON-OVERLAP OBSERVATIONS ONLY

STATE	STRATA	IDENT	SURVEY	MPRESHOG	MRESPHOG	MEXPFCTR	LHOGTOTL	EXPDHGTL
GEORGIA (CUTOFF=25000) (Cont'd)								
13	**40**	**13-88-15A**	**Q288**	**0**	**1**	**185.90**	**1305**	**27674**
13	**40**	**13-88-15B**	**Q388**	**0**	**5**	**309.83**	**1305**	**46122**
13	**40**	**13-88-15C**	**Q488**	**0**	**5**	**309.83**	**1105**	**39054**
13	**40**	**13-88-15D**	**Q189**	**0**	**4**	**309.83**	**1351**	**47748**
13	40	13-88-16A	Q288	0	1	185.90	584	33632
13	40	13-88-16B	Q388	0	3	309.83	697	66899
13	40	13-88-16C	Q488	0	3	309.83	677	64980
13	40	13-88-16D	Q189	0	4	309.83	862	82737
IDAHO (CUTOFF=5000)								
16	15	16-88-17A	Q288	0	1	93.60	64	5990
16	**15**	**16-88-18A**	**Q288**	**0**	**2**	**46.80**	**10**	**351**
16	**15**	**16-88-18B**	**Q388**	**0**	**2**	**78.00**	**40**	**2340**
16	**15**	**16-88-18C**	**Q488**	**0**	**2**	**97.50**	**47**	**3437**
16	**15**	**16-88-18D**	**Q189**	**0**	**2**	**97.50**	**112**	**8190**
16	22	16-88-19A	Q288	0	4	43.20	0	0
16	22	16-88-19B	Q388	0	4	72.00	88	5069
16	22	16-88-19C	Q488	3	2		.	.
16	22	16-88-19D	Q189	3	6		.	.
16	**31**	**16-88-20A**	**Q288**	**0**	**4**	**350.80**	**20**	**7016**
16	**31**	**16-88-20B**	**Q388**	**0**	**2**	**584.60**	**1**	**585**
16	31	16-88-21A	Q288	0	2	350.80	43	15084
16	31	16-88-21B	Q388	0	2	584.60	67	39168
16	31	16-88-21C	Q488	3	2	584.67	0	0
ILLINOIS (CUTOFF=50000)								
17	11	17-88-22A	Q288	0	1	173.00	2336	39591
17	11	17-88-22B	Q388	0	4	173.00	2820	47794
17	11	17-88-22C	Q488	0	2	173.00	3290	55760
17	11	17-88-22D	Q189	0	4	173.00	3158	53523
17	**12**	**17-88-23A**	**Q288**	**0**	**4**	**200.60**	**402**	**80641**
17	**12**	**17-88-23B**	**Q388**	**0**	**4**	**200.60**	**404**	**81042**
17	**12**	**17-88-23C**	**Q488**	**3**	**2**	**200.60**	**0**	**0**
17	**12**	**17=88-23D**	**Q189**	**0**	**2**	**407.89**	**257**	**104828**
17	12	17-88-24A	Q288	0	1	200.60	600	120360
17	12	17-88-24B	Q388	0	2	200.60	560	112336
17	12	17-88-24C	Q488	3	2	200.60	0	0
17	**12**	**17-88-25A**	**Q288**	**0**	**1**	**200.60**	**754**	**54169**
17	**12**	**17-88-25B**	**Q388**	**0**	**4**	**200.60**	**752**	**54026**
17	**12**	**17-88-25C**	**Q488**	**0**	**4**	**200.60**	**1004**	**72130**
17	**12**	**17-88-25D**	**Q189**	**0**	**5**	**200.60**	**954**	**68538**
17	20	17-88-26A	Q288	0	1	241.60	595	38702
17	20	17-88-26B	Q388	0	2	241.60	885	57566
17	20	17-88-26C	Q488	0	2	241.60	934	60753
17	20	17-88-26D	Q189	0	2	241.60	707	45988

INDIVIDUAL NOL OPERATION RESPONSES FOR
QUARTERLY AG SURVEYS FROM JUNE 1988 THROUGH MARCH 1989
NON-OVERLAP OBSERVATIONS ONLY

STATE	STRATA	IDENT	SURVEY	MPRESHOG	MRESHOG	MEXPFCTR	LHOGTOTL	EXPDHGTL
ILLINOIS (CUTOFF=50000) (Cont'd)								
17	40	17-88-27A	Q288	0	1	189.00	1200	61584
17	40	17-88-27B	Q388	3	5	189.00	0	0
17	40	17-88-27C	Q488	0	2	189.00	921	47266
17	40	17-88-27D	Q189	0	2	189.00	1200	61584
MICHIGAN (CUTOFF=15000)								
26	11	26-88-28A	Q288	1	6	169.40	4	678
26	11	26-88-28B	Q388	0	4	169.40	170	28798
26	**11**	**26-88-29A**	**Q288**	**0**	**1**	**169.40**	**205**	**7361**
26	**11**	**26-88-29B**	**Q388**	**0**	**2**	**169.40**	**90**	**3231**
26	**11**	**26-88-29C**	**Q488**	**0**	**2**	**338.80**	**268**	**19245**
26	**11**	**26-88-29D**	**Q189**	**0**	**4**	**338.80**	**273**	**19604**
26	11	26-88-30A	Q288	0	4	169.40	320	53480
26	11	26-88-30B	Q388	0	2	169.40	285	47631
26	11	26-88-30C	Q488	0	2	169.40	240	40110
26	11	26-88-31D	Q189	0	2	169.40	175	29247
26	**12**	**26-88-32A**	**Q288**	**0**	**1**	**191.40**	**131**	**16737**
26	**12**	**26-88632B**	**Q388**	**3**	**2**		**.**	**.**
26	**12**	**26-88-32C**	**Q488**	**3**	**6**		**.**	**.**
26	12	26-88-33A	Q288	0	1	191.40	122	7967
26	12	26-88-33B	Q388	0	5	191.40	200	13060
26	12	26-88-33C	Q488	0	2	319.00	125	13604
26	12	26-88-33D	Q189	0	2	319.00	146	15890
26	**20**	**26-88-34A**	**Q288**	**0**	**1**	**220.80**	**403**	**20868**
26	**20**	**26-88-34B**	**Q388**	**0**	**2**	**220.80**	**388**	**20091**
26	**20**	**26-88-34C**	**Q488**	**0**	**2**	**220.80**	**263**	**13618**
26	**20**	**26-88-34D**	**Q189**	**0**	**2**	**220.80**	**212**	**10977**
26	31	26-88-35A	Q288	2	6	299.30	0	0
26	31	26-88-35B	Q388	0	2	299.30	111	33048
26	31	26-88-35C	Q488	3	5	561.19	0	0
26	31	26-88-35D	Q189	0	2	561.19	152	84854

QUARTERLY AG SURVEYS FROM JUNE 1989 THROUGH MARCH 1990
NON-OVERLAP OBSERVATIONS ONLY

COLORADO (CUTOFF=10000)								
8	13	8-89-36A	Q289	0	1	96.40	376	23409
GEORGIA (CUTOFF=25000)								
13	20	13-89-37A	Q289	0	1	167.90	302	39451
13	20	13-89-37B	Q389	0	3	279.83	70	15240
13	20	13-89-37C	Q489	0	3	279.83	74	16111
13	20	13-89-37D	Q190	0	4	279.83	300	65316

INDIVIDUAL NOL OPERATION RESPONSES FOR
QUARTERLY AG SURVEYS FROM JUNE 1989 THROUGH MARCH 1990
NON-OVERLAP OBSERVATIONS ONLY

STATE	STRATA	IDENT	SURVEY	MPRESHOG	MRESPHOG	MEXPFCTR	LHOGTOTL	EXPDHGTL
GEORGIA (CUTOFF=25000) (Cont'd)								
13	20	13-89-38A	Q289	0	.	167.90	0	0
13	20	13-89-38C	Q489	0	1	5923.14	11	60501
13	20	13-89-38D	Q190	0	3	5923.14	22	121001
13	**20**	**13-89-39A**	**Q289**	**0**	**1**	**168.30**	**424**	**35680**
13	**20**	**13-89-39B**	**Q389**	**0**	**3**	**280.50**	**422**	**59185**
13	**20**	**13-89-39C**	**Q489**	**0**	**5**	**280.50**	**424**	**59466**
13	**20**	**13-89-39D**	**Q190**	**0**	**5**	**280.50**	**424**	**59466**
13	40	13-89-40A	Q289	0	1	185.90	654	59868
13	40	13-89-40B	Q389	0	2	309.83	3157	481657
13	40	13-89-40C	Q489	0	3	309.83	4680	714017
13	40	13-89-40D	Q190	0	3	309.83	5986	913271
13	**40**	**13-89-41A**	**Q289**	**0**	**1**	**185.90**	**1113**	**110057**
IDAHO (CUTOFF=5000)								
16	13	16-89-62A	Q289	0	2	69.60	0	0
16	13	16-89-42B	Q389	0	2	474.09	27	12747
16	13	16-89-42C	Q489	3	2	147.23	0	0
16	13	16-89-42D	Q190	0	2	147.23	42	6158
16	**15**	**16-89-43A**	**Q289**	**0**	**2**	**46.80**	**184**	**7750**
16	**15**	**16-89-43B**	**Q389**	**3**	**2**	**78.00**	**0**	**0**
16	**15**	**16-89-43C**	**Q489**	**0**	**2**	**78.00**	**25**	**1755**
16	**15**	**16-89-43D**	**Q190**	**3**	**2**	**78.00**	**0**	**0**
16	31	16-89-44A	Q289	0	2	350.80	52	18242
16	31	16-89-44B	Q389	0	5	584.67	64	37419
16	31	16-89-44C	Q489	0	2	584.67	72	42096
16	31	16-89-44D	Q190	0	2	584.67	76	44435
ILLINOIS (CUTOFF=50000)								
17	11	17-89-45A	Q289	0	1	173.00	3186	55328
17	11	17-89-45B	Q389	0	4	173.00	3359	58332
17	11	17-89-45C	Q489	0	4	173.00	3325	57741
17	11	17-89-45D	Q190	0	4	173.00	3742	64983
17	**11**	**17-89-46A**	**Q289**	**0**	**2**	**173.00**	**268**	**46364**
17	**11**	**17-89-46B**	**Q389**	**0**	**2**	**173.00**	**270**	**46710**
17	**11**	**17-89-46C**	**Q489**	**0**	**2**	**173.00**	**268**	**46364**
17	**11**	**17-89-46D**	**Q190**	**0**	**2**	**173.00**	**303**	**52419**
17	12	17-89-47A	Q289	1	5	200.60	257	51554
17	12	17-89-47B	Q389	0	5	200.60	257	51554
17	12	17-89-47C	Q489	0	2	200.60	288	57773
17	12	17-89-47D	Q190	0	5	200.60	288	57773
17	**12**	**17-89-48A**	**Q289**	**0**	**1**	**200.60**	**525**	**75225**
17	**12**	**17-89-48B**	**Q389**	**0**	**2**	**200.60**	**775**	**111046**
17	**12**	**17-89-48C**	**Q489**	**0**	**2**	**200.60**	**560**	**80240**
17	**12**	**17-89-48D**	**Q190**	**3**	**2**	**200.60**	**0**	**0**

INDIVIDUAL NOL OPERATION RESPONSES FOR
QUARTERLY AG SURVEYS FROM JUNE 1989 THROUGH MARCH 1990
NON-OVERLAP OBSERVATIONS ONLY

STATE	STRATA	IDENT	SURVEY	MPRESHOG	MREHOG	MEXPFCTR	LHOGTL	EXPDHGTL
ILLINOIS (CUTOFF=50000) (Cont'd)								
17	12	17-89-49A	Q289	0	4	200.60	1080	63060
17	12	17-89-49B	Q389	0	4	200.60	909	53076
17	12	17-89-49C	Q489	0	2	200.60	853	49806
17	12	17-89-49D	Q190	0	2	200.60	803	46886
17	**12**	**17-89-50A**	**Q289**	**0**	**1**	**200.60**	**85**	**17027**
17	**12**	**17-89-50B**	**Q389**	**0**	**4**	**200.60**	**123**	**24640**
17	**12**	**17-89-50C**	**Q489**	**0**	**2**	**401.20**	**147**	**58895**
17	**12**	**17-89-50D**	**Q190**	**0**	**2**	**401.20**	**121**	**48478**
MICHIGAN (CUTOFF=15000								
26	11	26-89-51A	Q289	0	1	169.40	118	19849
26	11	26-89-51B	Q389	0	4	282.33	116	32521
26	11	26-89-51C	Q489	0	1	282.33	150	42053
26	11	26-89-51D	Q190	0	1	282.33	156	43735
26	**11**	**26-89-52A**	**Q289**	**0**	**1**	**169.40**	**288**	**17904**
26	12	26-89-53A	Q289	0	1	191.40	0	0
26	12	26-89-53B	Q389	0	2	319.00	50	14178
26	12	26-89-53C	Q489	0	2	319.00	80	22684
26	12	26-89-53D	Q190	3	2	319.00	0	0
26	**12**	**26-89-54A**	**Q289**	**1**	**5**	**191.40**	**47**	**3069**
26	**12**	**26-89-54B**	**Q389**	**0**	**2**	**319.00**	**102**	**11101**
26	**12**	**26-89-54C**	**Q489**	**0**	**5**	**638.00**	**117**	**25467**
26	**12**	**26-89-54D**	**Q190**	**0**	**5**	**638.00**	**116**	**25250**
26	20	26-89-55A	Q289	0	1	220.80	200	18534
26	**40**	**26-89-56A**	**Q289**	**0**	**1**	**353.20**	**14**	**4945**
26	**40**	**26-89-56B**	**Q389**	**0**	**2**	**588.67**	**41**	**24135**
26	**40**	**26-89-56C**	**Q489**	**0**	**2**	**1177.33**	**43**	**50625**
26	**40**	**26-89-56D**	**Q190**	**0**	**4**	**1177.33**	**28**	**32965**

QUARTERLY AG SURVEYS FROM JUNE 1990 THROUGH DECEMBER 1990
NON-OVERLAP OBSERVATIONS ONLY

COLORADO (CUTOFF=10000)								
8	34	8-90-57A	Q290	0	4	63.80	143	3146
8	34	8-90-57B	Q390	0	2	106.33	184	6746
8	34	8-90-57C	Q490	0	2	106.33	286	10486
GEORGIA (CUTOFF=25000)								
13	20	13-90-58A	Q290	1	5	167.90	116	5791
13	20	13-90-58B	Q390	0	2	279.83	332	27625
13	20	13-90-58C	Q490	3	2	279.83	0	0

INDIVIDUAL NOL OPERATION RESPONSES FOR
QUARTERLY AG SURVEYS FROM JUNE 1990 THROUGH DECEMBER 1990
NON-OVERLAP OBSERVATIONS ONLY

STATE	STRATA	IDENT	SURVEY	MPRESHOG	MRESHOG	MEXPFCTR	LHOGTOTL	EXPDHGTL
GEORGIA (CUTOFF=25000) (Cont'd)								
13	**20**	**13-90-58A**	**Q290**	**0**	**1**	**167.90**	**440**	**32513**
13	**20**	**13-90-58B**	**Q390**	**0**	**2**	**279.83**	**370**	**45566**
13	**20**	**13-90-58C**	**Q490**	**0**	**2**	**279.83**	**470**	**57882**
13	20	13-90-59A	Q290	0	1	167.90	1225	28281
13	20	13-90-59B	Q390	0	4	279.83	1225	47134
13	20	13-90-59C	Q490	0	4	279.83	710	27318
13	**40**	**13-90-60A**	**Q290**	**0**	**1**	**185.90**	**70**	**6092**
13	**40**	**13-90-60B**	**Q390**	**0**	**2**	**309.83**	**84**	**12183**
13	**40**	**13-90-60C**	**Q490**	**0**	**2**	**929.50**	**58**	**25237**
13	40	13-90-61A	Q290	0	1	185.90	820	41879
IDAHO (CUTOFF=5000)								
16	15	16-90-62A	Q290	0	1	46.80	35	1638
16	15	16-90-62B	Q390	0	2	78.00	89	6942
16	15	16-90-62C	Q490	0	2	78.00	2	156
16	**31**	**16-90-63A**	**Q290**	**0**	**1**	**350.80**	**38**	**13330**
ILLINOIS (CUTOFF=50000)								
17	11	17-90-64A	Q290	0	1	115.33	405	22663
17	11	17-90-64B	Q390	0	4	192.22	390	36373
17	11	17-90-64C	Q490	0	2	480.55	675	157385
17	**11**	**17-90-65A**	**Q290**	**1**	**5**	**115.33**	**2070**	**66646**
17	**11**	**17-90-65B**	**Q390**	**0**	**4**	**192.22**	**755**	**40514**
17	**11**	**17-90-65C**	**Q490**	**0**	**5**	**192.22**	**755**	**40514**
17	11	17-90-66A	Q290	2	1	115.33	0	0
17	11	17-90-66B	Q390	3	5	192.22	0	0
17	11	17-90-66C	Q490	0	5	384.44	405	155698
17	**11**	**17-90-67A**	**Q290**	**0**	**1**	**115.33**	**237**	**26667**
17	**11**	**17-90-67B**	**Q390**	**0**	**4**	**192.22**	**261**	**48946**
17	**11**	**17-90-67C**	**Q490**	**0**	**4**	**230.67**	**247**	**55586**
17	12	17-90-68A	Q290	0	2	200.60	30	6018
17	12	17-90-68B	Q390	3	4	334.33	0	0
17	12	17-90-68C	Q490	0	2	334.33	183	61182
17	**12**	**17-90-69A**	**Q290**	**0**	**1**	**200.60**	**852**	**60475**
17	12	17-90-70A	Q290	0	1	200.60	478	32103
17	12	17-90-70B	Q390	0	2	334.33	387	43319
17	12	17-90-70C	Q490	0	2	334.33	526	58877
17	**12**	**17-90-71A**	**Q290**	**0**	**1**	**200.60**	**348**	**51193**
17	**12**	**17-90-71B**	**Q390**	**0**	**4**	**334.33**	**333**	**81643**
17	**12**	**17-90-71C**	**Q490**	**0**	**5**	**334.33**	**333**	**81643**
17	20	17-90-72A	Q290	0	1	241.60	1192	113866
17	20	17-90-72B	Q390	0	5	402.67	1192	189779
17	20	17-90-72C	Q490	0	2	402.67	1412	224805

INDIVIDUAL NOL OPERATION RESPONSES FOR
QUARTERLY AG SURVEYS FROM JUNE 1990 THROUGH DECEMBER 1990
NON-OVERLAP OBSERVATIONS ONLY

STATE	STRATA	IDENT	SURVEY	MPRESHOG	MRESPHOG	MEXPFCTR	LHOGTOTL	EXPDHGTTL
ILLINOIS (CUTOFF=50000) (Cont'd)								
17	20	17-90-73A	Q290	0	1	241.60	238	15682
17	20	17-90-73B	Q390	0	2	402.67	468	51395
17	20	17-90-73C	Q490	0	2	402.67	381	41841
MICHIGAN (CUTOFF=15000)								
26	11	26-90-74A	Q290	0	4	96.70	303	19052
26	**12**	**26-90-75A**	**Q290**	**0**	**1**	**99.20**	**131**	**12995**
26	**12**	**26-90-75B**	**Q390**	**0**	**2**	**165.33**	**122**	**20170**
26	**12**	**26-90-75C**	**Q490**	**0**	**2**	**165.33**	**150**	**24799**
26	20	26-90-76A	Q290	0	1	154.80	1072	49593

CODE DEFINITIONS VALID FOR THE 1987 FRAME ONLY

MPRESHOG
(HOG PRESENCE STATUS)

1 - POSITIVE FOR HOGS BUT INACCESSIBLE OR REFUSAL
2 - HOG STATUS UNKNOWN
3 - ZERO HOGS - INCLUDES BOTH USABLES & UNUSABLES
4 - DATA COMPLETE - INCLUDES BOTH POSITIVES & ZEROS

MRESPHOG
(SURVEY RESPONSE STATUS)

1 - MAIL RESPONSE - USABLE
2 - TELEPHONE (CATI OR PERSONAL) RESPONSE - USABLE
3 - PERSONAL INTERVIEW RESPONSE - USABLE
4 - HOG NUMBER ESTIMATED - USABLE
5 - KNOWN ZERO HOGS - USABLE
6 - MAIL REFUSAL - UNUSABLE
7 - TELEPHONE REFUSAL - UNUSABLE
8 - PERSONAL INTERVIEW REFUSAL - UNUSABLE
9 - INACCESSIBLE

CODE DEFINITIONS FOR THE 1988, 1989 AND 1990 FRAMES

MPRESHOG
(HOG PRESENCE STATUS)

0 - POSITIVE FOR HOGS WITH DATA - USABLE
1 - POSITIVE FOR HOGS WITH NO DATA - UNUSABLE
2 - UNKNOWN HOG STATUS WITH NO DATA - UNUSABLE
3 - KNOWN ZERO - USABLE

MRESPHOG
(RESPONSE STATUS)

1 - MAIL RESPONSE - USABLE
2 - PERSONAL TELEPHONE RESPONSE - USABLE
3 - CATI TELEPHONE RESPONSE - USABLE
4 - PERSONAL INTERVIEW RESPONSE - USABLE
5 - HOG NUMBER ESTIMATED - USABLE
6 - KNOWN ZERO HOGS - USABLE
7 - MAIL REFUSAL - UNUSABLE
8 - TELEPHONE REFUSAL - UNUSABLE
9 - CATI REFUSAL - UNUSABLE
10 - PERSONAL INTERVIEW REFUSAL - UNUSABLE
11 - INACCESSIBLE - UNUSABLE

GENERAL AREA STRATA DEFINITIONS

(% UNDER CULTIVATION)

11 - 75% OR MORE
12 to 19 - 50 to 75%
20 to 29 - 15 to 49%
30 to 39 - Ag/Urban
40 to 49 - less than 15%
50 - Non-Agricultural

APPENDIX F

SUMMARY OUTLIER CHARACTERISTICS BY STATE

	COLORADO (14 Surveys)	GEORGIA (15 Surveys)	IDAHO (13 Surveys)	ILLINOIS (15 Surveys)	MICHIGAN (14 Surveys)
OUTLIER CUTOFF	(10,000)	(25,000)	(5,000)	(80,000)	(15,000)
OUTLIER OCCURRENCES					
NUMBER OF:					
LIST					
Unique Operations	5	3	3	18	67
Total Occurrences	8	3	4	19	87
% of Survey Outlier Total	53%	8%	20%	26%	74%
NOL					
Unique Operations	6	15	11	26	18
Total Occurrences	7	36	16	54	31
% of Survey Outlier Total	47%	92%	80%	74%	26%
COMBINED LIST/NOL					
Unique Operations	11	18	14	44	85
Total Occurrences	15	39	20	73	118
AVERAGE OF:					
LIST					
Unique Operations	0.36	0.20	0.23	1.20	4.79
Total Occurrences	0.57	0.20	0.31	1.27	6.21
NOL					
Unique Operations	0.43	1.00	0.85	1.73	1.29
Total Occurrences	0.50	2.40	1.23	3.60	2.21
COMBINED LIST/NOL					
Unique Operations	0.79	1.20	1.08	1.93	5.98
Total Occurrences	1.07	2.60	1.54	4.87	8.41
OUTLIER DIRECT EXPANSION					
AVERAGE VALUE PER OCCURRENCE					
LIST	12,235	27,669	8,849	82,969	21,782
NOL	17,551	108,214	17,309	86,790	30,035
COMBINED LIST/NOL	14,716	102,019	15,617	85,782	23,950
AVERAGE TOTAL PER SURVEY					
LIST	6,991	5,533	2,723	105,094	135,361
NOL	8,776	259,715	21,304	306,659	66,505
COMBINED LIST/NOL	15,767	265,248	24.027	411,753	201,866
AVERAGE PERCENTAGE of SURVEY					
LIST	3.0%	0.4%	3.0%	2.0%	11.0%
NOL	4.0%	21.0%	24.0%	6.0%	5.0%
COMBINED LIST/NOL	7.0%	21.4%	27.0%	8.0%	16.0%
AVERAGE SURVEY TOTAL	238,733	1,244,056	89,561	5,444,381	1,234,773

☆U.S. Government Printing Office : 1992 - 311-355/60165